Development of crystallographic techniques and their application to several protein targets

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by Sam Horrell

July 2015

Abstract

Since its first use to solve the structure of sodium chloride in 1915 X-ray crystallography has developed significantly to become the premier technique for obtaining 3D structural information of small molecules and macromolecules alike. As the technique continues to develop and focus its attention on weak diffraction from the likes of micro-crystals and poorly packed crystals of membrane proteins and large protein complexes; as well as ultra-high resolution data and weak anomalous signal from native atoms, data quality is becoming more and more important. Data quality is particularly important in the wake of long wavelength macromolecular crystallography (MX) for phasing using anomalous signal from native sulphur and phosphorous atoms in proteins and DNA.

This thesis first investigated the use of a new sample handling technique using a humidity controlled stream to preserve macromolecular crystals while excess surrounding solvent is removed (Chapter 2). Following the successful development of this technique the effects of excess surrounding solvent on data quality was assessed when collecting at standard MX X-ray wavelengths (~ 1 Å) and longer X-ray wavelengths (~ 2 Å). Datasets were collected from large populations of control and test crystals at standard and longer wavelengths to allow robust statistical methods to be applied; a practice not widely adopted in method development studies in X-ray crystallography. This made it possible to assess the small differences in data quality in the presence and absence of excess surrounding solvent. The effects of surrounding solvent at longer wavelengths appear to be protein dependent with some proteins tested showing no significant difference and others a significant decrease in data quality at longer wavelengths (Chapter 3).

Originally this project aimed to use the new long wavelength in-vacuum MX beamline, I23, at Diamond Light Source UK to carry out phasing experiments using native sulphurs for structure solution. However, the considerable complexity involved in developing in-vacuum MX meant these experiments could not be carried out during the time frame of this thesis. Chapters 4 and 5 outline the production of a novel cancer protein (cancerous inhibitor of protein phosphatase 2A) and two

protein targets from the *Achromobacter xylosoxidans* (*Ax*) genome intended for sulphur single wavelength anomalous dispersion phasing experiments on I23. Of these proteins the structure of *Ax*-α/β hydrolase was solved by conventional methods, the structure of which is discussed in Chapter 5.

Of the protein crystals used in long wavelength data quality experiments in Chapter 3 the molecular biology of *PA*3825-EAL, a biofilm regulating protein essential to the swarming ability of *Pseudomonas aeruginosa*, was investigated further. The crystal structure of *PA*3825-EAL was solved in the resting, substrate bound and product bound states to high resolution. Comparison of the crystal structures of monomeric and dimeric *PA*3825-EAL with the inactive dimeric structure of MucR-EAL suggests dimerisation via helix 8 plays a role in inhibition of EAL domains. Prior to this, dimerisation was thought to be an activating factor in EAL domains. The product bound state of *PA*3825-EAL showed the presence of a previously unreported third metal binding site which may form an essential component of the reaction mechanism of EAL domains. Inability of MucR-EAL to incorporate this third metal due to dimerisation may explain the lack of activity despite possessing the conserved catalytic residues necessary.

The fast detector technology and improvements in automated data processing software that allowed diffraction data for large populations of crystals to be collected in Chapters 2 and 3 have also been applied to development of a serial data collection technique. Of 159 datasets collected from 8 crystals of a copper nitrite reductase from *Achromobacter cycloclastes*, 45 datasets from a single crystal were analysed to observe the reaction mechanism using high resolution crystal structures. X-ray radiolysis initiated the reaction and high resolution data allowed the conversion of nitrite (NO_2) to nitric oxide (NO) to be observed in the crystal. Other aspects of the reaction were investigated from the data series including a conserved water chain connecting the copper sites which may act as a proton wire to donate a proton and produce NO. This technique may have wide applications to the study of the reaction mechanisms of other metallo-proteins.

Contents

Tables

Figures

Abbreviations

ABH	α/β hydrolase
Ac	Achromobacter cycloclastes
*Ac*NiR	*Achromobacter cycloclastes* nitrite reductase
APC	Adenomatous polyposis coli
Arm	Armadillo repeat
Ax	*Achromobacter xylosoxidans*
*Ax*ABH	*Achromobacter xylosoxidans* α/β hydrolase
*Ax*UDT	*Achromobacter xylosoxidans* UTP-glucose-1-phosphate uridyltransferase
bdlA	Biofilm dispersion locus
BLAST	Basic Local Alignment Search Tool
BLUF	Blue-light using FAD
BPM2	Bacteriocin pectocin M2
BSM	Bacteriocin syringacin M
cd1-NiR	Cytochrome cd1 nitrite reductase
CDC	Centre for disease control
CdG	Cyclic di-GMP
CIP2A	Cancerous inhibitor of protein phosphatase 2A
CIP2A-C	CIP2A C-terminal domain
CIP2A-N	CIP2A N-terminal domain
CUP	Chaperone/usher protein
CV	Column volume
DGC	Diguanylate cyclase
DUF	Domain of unknown function
FAP	Functional amyloids in *Pseudomonas*
FMS	Free-mounting system
GST	Glutathione S-transferase

HC1	Humidity control device
HC-T	Humidity control transferred
HEWL	Hen egg white lysozyme
HRL	High resolution limit
HTH	Helix turn helix
IP (site)	Primary inhibition site
keV	Kilo-electron volts
MAD	Multi-wavelength anomalous dispersion
MALS	Multi-angle light scattering
MBP	Maltose binding protein
MGy	Mega grey
MIR	Multiple isomorphous replacement
MR	Molecular replacement
MX	Macromolecular crystallography
PDB	Protein data bank
NLS	Nuclear localisation sequence
PAD	PP2A activating drug
PDE	Phosphodiesterase
PEEK	polyetheretherketone
PEG	Polyethylene glycol
pGpG	5’- Phosphoguanylyl- (3’ –> 5’)- guanosine
PP2A	Protein phosphatase 2A
PULSA	Pulsed UV laser soft ablation
R_{mease}	Redundancy independent merging factor
R_{merge}	Linear merging factor
$R_{p.i.m.}$	Precision-indicating merging factor
SAD	Single-wavelength anomalous dispersion

S-SAD	Sulphur single-wavelength anomalous dispersion
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide agarose gel electrophoresis
SEC	Size exclusion chromatography
siRNA	Small interfering RNA
SUMO	Small ubiquitin-related modifier
T1Cu	Type one copper
T2Cu	Type two copper
TIM	Triosephosphate isomerase
UTR	Untranslated region
XAS	X-ray absorption spectroscopy

Acknowledgements

I would like to thank the following;

My supervisors Dr Richard Strange, Dr Armin Wagner and Prof. Samar Hasnain for their support, guidance, patience and advice throughout my studies.

Dr Domenico Bellini for all lessons in high throughput protein production and crystallisation. As well as occasional trips to the pub.

The I23 team, particularly Dr Ramona Duman for all her help with the HC1 project and Dr Vitaly Mykhaylyk for always making Friday meetings more interesting.

Dr Svetlana Antonyuk for her advice on all things AcNiR.

The staff of the Diamond Light Source and Research Complex at Harwell for providing and maintaining excellent facilities for scientific research during my stay. Particularly Dr Juan Sanchez-Weatherby for help with the HC1 and Maria Rosa for maintaining the labs and making many litres of media for me.

Dr Rhys Grinter for providing protein samples for HC1 tests.

The students of the Molecular Biophysics group and the members of the Oxford Ultimate Club for the lighter side of life.

The BBSRC and Diamond Light Source for providing a 4 year PhD studentship award supporting me through the course of my studies.

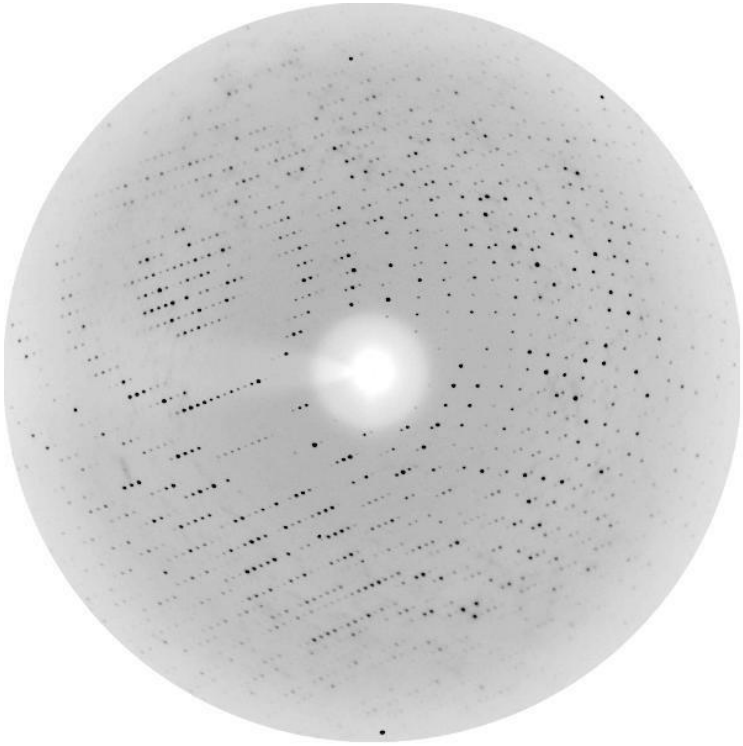
My parents and brother for their continued love and support throughout this process.

Preface

This Thesis presents original research carried out by the author submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy to the University of Liverpool

This research consists of a number of studies using X-ray crystallography to (1) assess the effects of humid air streams, sample handling and excess surrounding solvent on data quality in long wavelength data collection using robust statistical analysis; (2) produce protein crystals with a view to perform in-vacuum S-SAD phasing experiments; (3) investigate the structure of a novel α/β hydrolase enzyme from the *Alcaligenes xylosoxidans* genome; (4) better understand the regulation of cyclic di-GMP cleavage by *PA*3825-EAL and other EAL domain containing proteins and (5) to develop the use of fast detectors and automatic processing pipelines for serial data collection of metallo-proteins to analyse their reaction mechanisms and produce structural movies.

Chapter 1 - X-ray Crystallography Theory



1.1. Introduction

X-ray crystallography is the premier technique for obtaining three dimensional structural information for chemical and biological molecules. The method involves measurement of the intensity of X-rays diffracted by a crystal made up of a regular repeating lattice of macromolecules or other small molecules. There are several textbooks that cover the principles of X-ray diffraction (Glusker and Trueblood, 1985; Blundell and Johnson, 1976; Blow, 2002 and Rupp, 2010) so will only be discussed briefly here.

When X-rays are scattered by electrons they are either scattered elastically, with no energy loss and X-ray photons are emitted at the same frequency as it entered, or inelastically, where electrons accept some energy from incoming X-ray photon reducing the energy of the emitted photon. When X-rays diffract from a crystal the contribution of the molecules in the regular lattice amplifies the signal you would see from a single molecule to produce discrete diffraction spots. Using Bragg's law (equation 1.1) we can quantify the angle at which X-rays are scattered (θ) relative to the distance between points in the crystal lattice (d_{hkl}), as shown in Figure 1.1 where X-rays are seen to be reflected from a set of planes in a crystal.

$$n\lambda = 2d_{hkl}\sin\theta$$

(1.1)

When X-rays scattered from a crystal satisfy the diffraction condition defined by Bragg's law we observe discrete diffraction spots. Fulfilment of the diffraction condition is elegantly depicted by superposition of the Ewald sphere with a reciprocal lattice where the lattice origin (000) is in line with the incoming X-ray beam (Fig. 1.2). Wherever a reciprocal lattice point intersects with the Ewald sphere the diffraction condition is fulfilled. By rotating the crystals we cause other reciprocal lattice points to intersect with the Ewald sphere allowing us to measure a complete diffraction dataset.

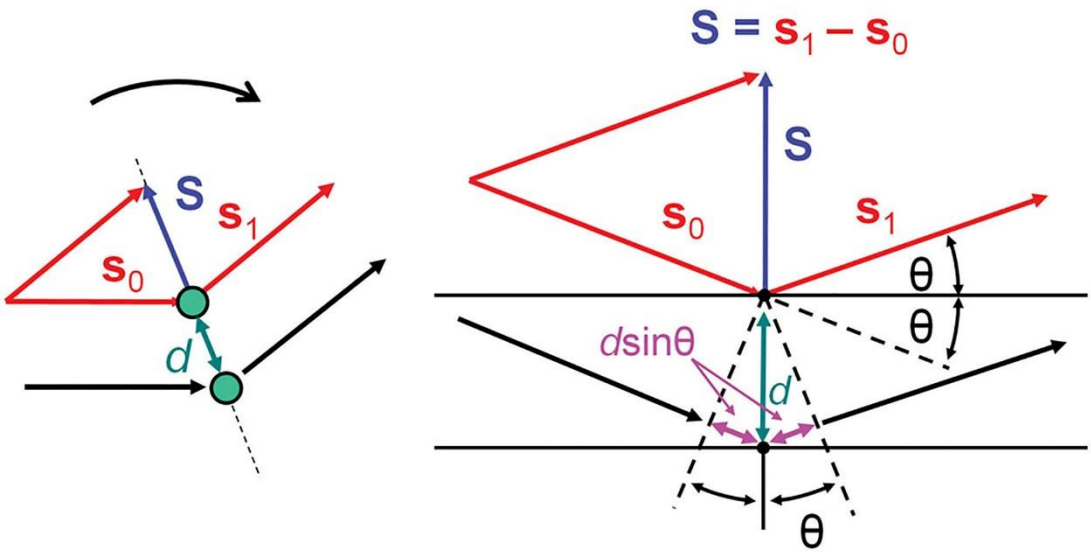


Fig. 1.1: A visual representation of Bragg's law with X-ray diffraction interpreted as a reflection form a lattice plane. This demonstrates the difference between diffraction planes to be 2d•Sinθ to fulfil the diffraction condition. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

Analysis of X-rays scattered by electrons allows us to calculate how the electrons are distributed through a crystal using the electron density equation, which expresses electron density at any point (x,y,z) in the unit cell as a Fourier series; a representation of a function as the sum of sine wave.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F(hkl)\exp(-2\pi i(hx+ky+lz))$$

(1.2)

The structure factors F(hkl) represent the information obtained through the crystallographic experiment which is defined as,

$$F(hkl) = \sum_{j=1}^{N} f_j \exp(2\pi i(hx_j + ky_j + lz_j))$$

(1.3)

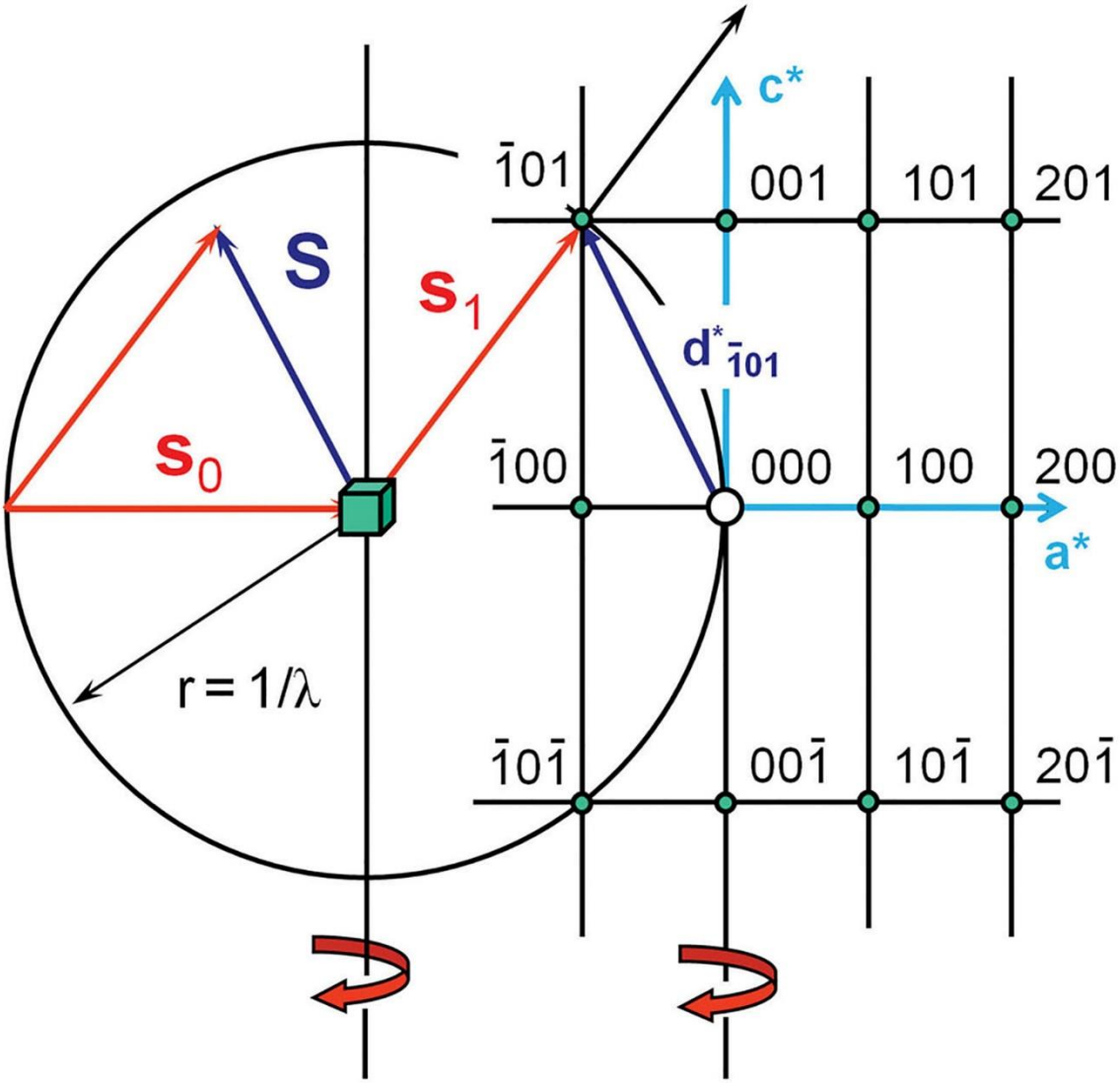


Fig. 1.2: The Ewald Sphere superposed with a reciprocal lattice. Reciprocal lattice points that intersect the Ewald Sphere fulfil the diffraction condition set out by Bragg's law. By rotating the crystal different points on the reciprocal lattice will fulfil the diffraction condition. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

where V represents the volume of the unit cell, h, k, and l represent the Miller indices (reciprocal index vectors representing 3D crystallographic lattice planes) of any particular reflection, xj, yj and zj represent the coordinates of the jth atom and fj is the scattering factor of the jth atom. The structure factors can be simplified to,

$$F(hkl) = F_{hkl} \exp(i\alpha_{hkl})$$

(1.4)

where F_{hkl} represents the measured amplitude obtained during crystallographic experiments as intensities and α_{hkl} is the phase of the structure factor for all measured reflections which must be determined mathematically to "solve" the structure of the crystallised molecule.

1.2. Data Collection

Diffraction data for all test crystals and structures solved in this thesis were collected at Diamond Light Source, UK, on beamlines I02, I03, I04, I04-1 and I24 on Dectris Pilatus 6M detectors. All data were collected using fine phi data collection strategies to minimise the contribution of background noise, allow 3D profile fitting of partial observations to reconstruct full reflections and reduce overlap in reflections from intersecting lunes (Mueller et al., 2012). To determine the amount of data required for a complete dataset the space group of the macromolecular crystal must be determined.

1.3. Space Groups

Space groups represent a group of mathematical symmetry opperations including translational, rotational (two-, three-, four- and six-fold operation) and translational-rotational operations (screw-axes and glide planes). Translation opperations result in fourteen Bravais lattices which give 230 unique space groups when combined with up to three other symmetry opperators. Due to the chirality of proteins and the need to maintain handedness in protein crystals, mirror planes, glide planes, improper rotations and inversion opperations are not allowed; leaving 65 chrial space groups in MX.

In most cases space groups are determined through the presence of systematic absences in diffraction data, these result from translational symmetry elements caused by Bravais translations and screw-axes which give reflections with zero intensity along specific directions in the reciprocal lattice. The exception to this are polar space groups with a single unique rotation/screw axis, such as *P6 or P*6_1, whose symmetry opperators do not fix the unit cell origin on a crystallographic axis. To

correctly identify the anantiomorph of these space groups an atomic position is required to fix the origin. Crystal structures may need to be reindexed if processed in the wrong anantiomorph.

1.4. Data reduction

Data reduction is a two-step process consisting of, (1) indexing diffraction spots to assign consistent unit cell vectors (h,k,l) to the diffraction pattern and (2) calculating reflection intensities. All data in this thesis were processed with XDS. XDS uses a robust auto-indexing system which derives reciprocal cell axis lengths from difference vectors between pairs of neighbouring reciprocal lattice vectors to produce a "reduced cell" in P1 and the lattice is fitted to the strongest 3000 reflections (Kabsch, 2010b, Kabsch, 2010a). The lattice is then refined against the strongest 3000 reflections followed by all found reflections and the Bravais lattice is calculated and fitted to possible space groups.

Integration involves accurate measurement of reflection intensities over multiple frames as all data collected in this thesis were collected using fine phi slicing so only partial reflections were recorded. XDS first calculates the frames and pixel positions which contribute to each reflection and generates a coordinate system. An average profile of strong reflections is then generated and the intensity is estimated by determining the background level and generating 3D profiles of contributing pixels for each reflection (Kabsch, 2010b, Kabsch, 2010a).

Individual images are then scaled to account for variations in the intensity of the beam, the illuminated crystal volume, absorption of the beam, radiation damage, detector sensitivity and differences in crystal size/order when scaling multiple crystals (Kabsch, 2010a). Correction factors are defined by,

$$\Psi_{hl} = (I_{hl} - g_l G_l I_h)/\sigma_{hl} \tag{1.5}$$

where h represents a unique reflection, hl represents symmetry-related reflections to h that are corrected by the reciprocal scaling factor g_l. I_{hl} is the reflection's and symmetry related reflection's weighted mean and σ_{hl} their standard deviation. Ih represents the "true" intensities as defined in detail in Kabsch (2010a). Finally data are merged into a single dataset and structure factors are calculated.

1.5. Data Quality

A number of metrics are applied to merged diffraction data to determine the relative error between measured reflections in individual resolution shells and the overall dataset. The signal-to-noise ratio, commonly referred to as the I/σ(I) is defined as,

$$\frac{|I|}{\sigma(I)} = \frac{1}{N}\sum_{h}^{N}\frac{I_{(h)}}{\sigma(I_{(h)})} \tag{1.6}$$

where N is the number of reflections in a given resolution shell and $I_{(h)}$ is the intensity of a given reflection. A variety of merging R values are commonly used to measure the error between individual reflections in defined resolution shell and over the whole dataset. As resolution increases reflections become weaker and relative error increases which is represented as an increase in these R values. The most common data quality indicator is the linear merging R factor (R_{merge}). However, R_{merge} fails to account for redundancy making low redundancy data appear better than high redundancy data (Diederichs and Karplus, 1997). As a result R_{merge} has been altered to give the redundancy independent merging R value (R_{meas}) with,

$$R_{meas} = \frac{\sum_h \left(\frac{N}{N-1}\right)^{\frac{1}{2}} \sum_{i=1}^{N} \left| I_{(ih)} - \bar{I}_{(h)} \right|}{\sum_h \sum_{i=1}^{N} I_{(ih)}}$$

(1.7)

where h is any given reflection, I_h is the intensity of an individual reflection, $\bar{I}_{(h)}$ is the average intensity of symmetry related reflections and N is the total number of reflections in the defined resolution bin. The precision-indicating merging factor ($R_{p.i.m.}$), which adds $\left(\frac{1}{N-1}\right)^{\frac{1}{2}}$ to the summation of reflection in the R_{merge} function in a similar vein to R_{meas}, has also been adopted as it shows increasing precision of intensity measurements with increasing redundancy (Weiss and Hilgenfeld, 1997, Weiss, 2001).

Another indicator becoming more common place is the use of correlation coefficients as a measure of data quality and maximum high resolution limit. Unmerged data is separated into two randomly selected half sets of unique reflections and calculate the correlation coefficient between averages half sets to give the $CC_{1/2}$ (Karplus and Diederichs, 2012). The correlation between the CC1/2 and the CC_{true} (the correlation of the averaged dataset with noise-free true signal), CC*, is calculated by,

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}}$$

(1.8)

1.6. Phasing

As defined earlier in equation 1.4, structure factors consist of a magnitude and a phase term; of which the magnitude is measured from the diffraction pattern while the phase information is lost. Determining the phase is essential to obtaining accurate structural information, as is elegantly shown in Kevin Cowtan's example of the Fourier Duck and Fourier Cat (Taylor, 2010); where

combining the magnitude from the Fourier transform of a duck (Fig. 1.3a) with the phase from the Fourier transform of a cat (Fig.1.3b) results in reconstruction of the image of the cat (Fig. 1.3c). There are several different methods that have been developed for determination of phases for macromolecular structures. The first used being multiple isomorphous replacement (MIR) by Perutz on the structure of haemoglobin using two bound mercury atoms (Green et al., 1954).

1.6.1. Multiple Isomorphous Replacement

Scatter from heavy atoms dominates the scattering pattern with scatter roughly proportional to atomic number. By introducing heavy atoms into a protein crystal we can introduce differences into to the diffraction pattern and the Patterson map (a 2D or 3D representation of all possible interatomic vectors calculated from intensities with a phase of zero) when compared with the native crystal. By subtracting the contribution of the native structure from the heavy atom structure we generate difference intensity data which represents the substructure of the heavy atoms. Using this much smaller number of atoms the heavy atom position can be determined using a difference Patterson map/direct methods to generate initial phases and electron density maps.

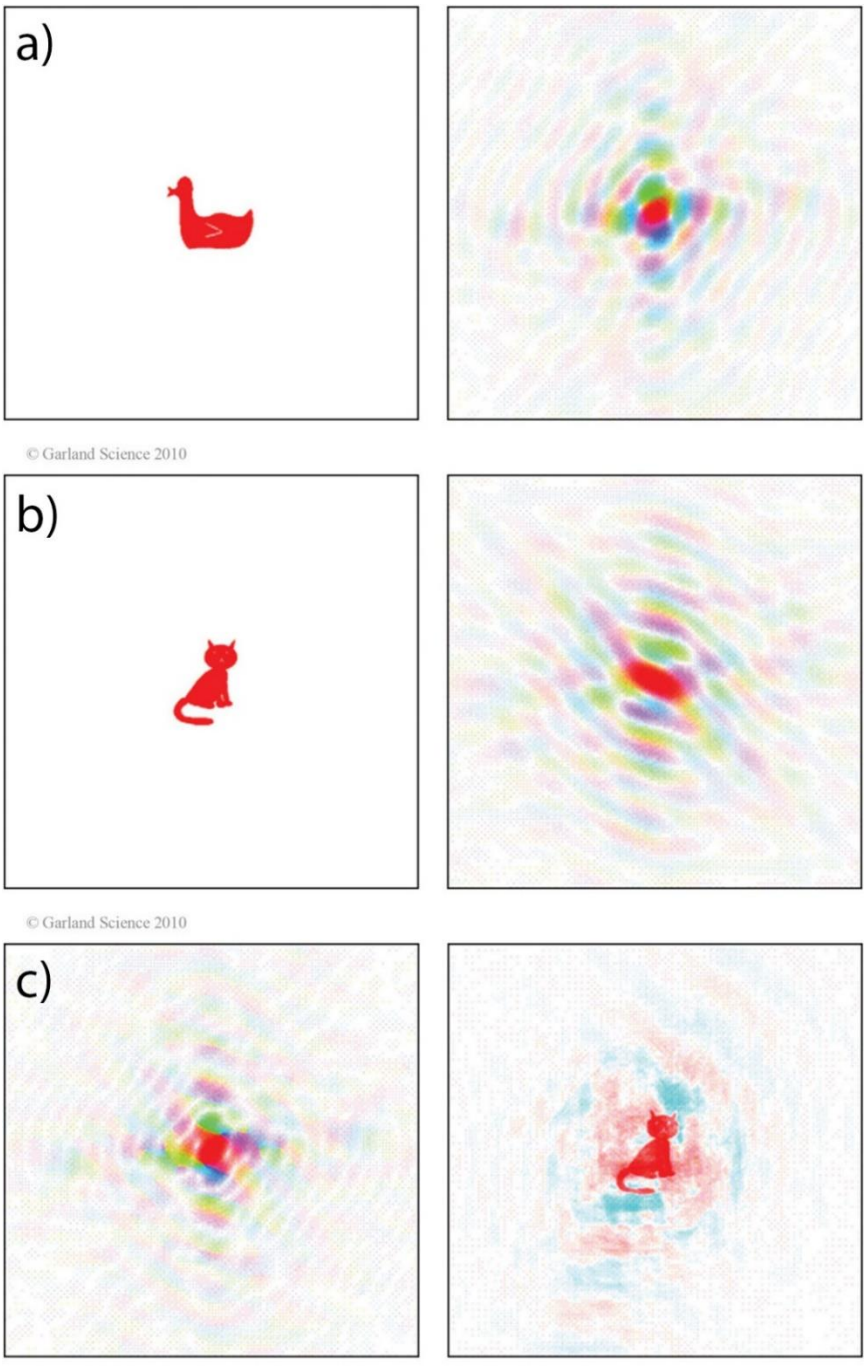


Fig. 1.3: The Importance of the Phase: a) The Fourier Duck (left) and its Fourier transform (right). b) The Fourier Cat (left) and its Fourier transform (right). c) A combination of the amplitude from the Fourier Duck and phase of the Fourier Cat (left) and the reverse Fourier transform (right). Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

1.6.2. Molecular Replacement

Molecular replacement (MR) involves the use of a known homologous structure, preferably with 30% or higher sequence identity and similar secondary structure, to determine the initial phases for an unknown structure. As of 2010 structures solved by MR account for over 95% of structures in the PDB (Vagin and Teplyakov, 2010b). Six parameters are applied to homologous structures, a 3D rotation search and a 3D translation search, to position the known structure on top of the unknown crystal structure and calculate initial phases. MR depends heavily on the Patterson function, which is used to calculate a vector map of a structure through a Fourier transform of intensities and amplitudes with a phase of zero. These maps represent the interatomic vectors of the structure which can be used to orientate a model in the unit cell of the unknown crystal structure. The self-Patterson of a known structure can be calculated in any orientation (rotation) and compared with the self-Patterson of the unknown structure. When these self-Pattersons match, the correct orientation of intramolecular parts of the structure have been determined and the fit is measured with the cross rotation function developed by Rossmann and Blow (1962),

$$R = \int_{U} P_2(X_2) P_1(X_1)\, dx_1 \tag{1.9}$$

where $P_1(X_1)$ represents the Patterson from the diffraction data, $P_2(X_2)$ the Patterson of the rotated model. The integration is carried out over the unit cell volume, U. Once the rotation function has been found the translation function orientates the structure with respect to the origin of the crystal. In a similar manner to the rotation function, the model is moved, a new Patterson generated and compared to the unknown structure for overlaps.

1.6.3. Phasing With Anomalous Signal

Much like the differences introduced into the diffraction pattern and Patterson map in multiple isomorphous replacement, anomalous signal can be utilised in a similar manner by altering the atomic scattering factor. The atomic scattering factor consists of three terms, (1) the normal scattering term (f^0) which is dependent on the Bragg angle; (2) the dispersive term (f') which affects the normal scattering factor and (3) the absorption term (f'') which causes a phase change of +90°. Both f' and f'' contribute to the anomalous scattering factor and are wavelength dependent; appearing at the absorption edge of an element when an electrons is ejected from a lower shell leaving a space for higher shell electrons to fill. Descending electrons produce X-ray emissions specific to absorbing elements which is measured as anomalous signal (Taylor, 2010).

$$f_{(\boldsymbol{s},\lambda)} = f^0_{(s)} + f'_{(\lambda)} + if''_{(\lambda)}$$

(1.10)

The presence of this anomalous signal causes a breakdown in Friedel's law ($|F_{hkl}| = |F_{\text{-}h\text{-}k\text{-}l}|$) as shown in Figure 1.4. Without anomalous signal Friedel's law holds with the sum of all normal and anomalous scattering atoms ($\mathbf{F_{PA}}$ and $\mathbf{F^{-}_{PA}}$) being symmetrical over the real axis (Fig. 1.4a). In the presence of anomalous signal ($\mathbf{F_A}$) the combination of the dispersive term (green) and absorption term (purple) (Fig. 1.4b) cause a significant difference between $\mathbf{F_{PA}}$ and $\mathbf{F^{-}_{PA}}$ (dark blue) over the real axis (Fig. 1.4c). Much like isomorphous replacement, this results in differences in the diffraction pattern and Patterson map which can be interpreted for phasing; without potentially introducing non-isomorphism with large disruptive heavy atoms.

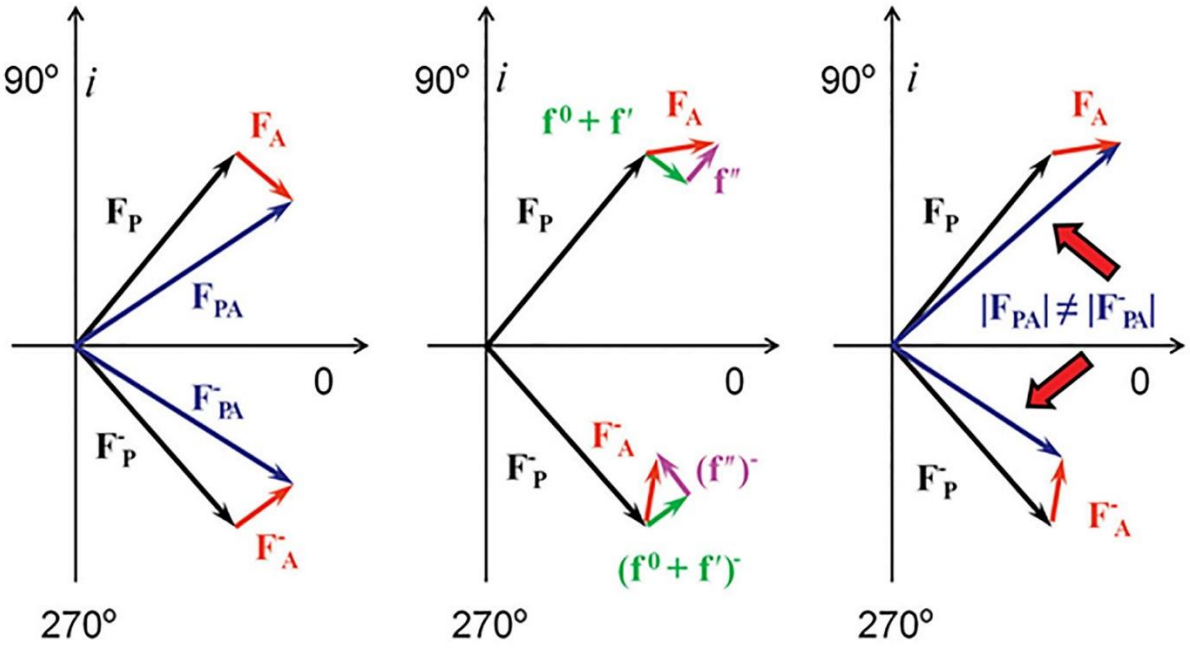


Fig. 1.4: Breaking Friedel's Law with anomalous signal. F_P represents the sum of non-anomalous scattering atoms (black), F_A the contribution of anomalous scattering atoms (red), $f^0 + f'$ are the real components of F_A (green) and f″ the imaginary component of F_A (purple). a) Visual representation of Friedel's law without anomalous signal where hkl and –h-k-l are mirrored across the real axis. b) The contribution of the real and imaginary components of anomalous scattering. c) The breakdown of Friedel's law where F_{PA} and F^-_{PA} are no longer mirrored across the real axis. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

1.6.4. MAD Phasing

Multiwavelength Anomalous dispersion (MAD) data are collected from a single crystal at several wavelengths to generate the strongest differences in anomalous data. Wavelengths used generally include the absorption peak where *f″* is at its maximum, the point of inflection where *f′* is at its minimum and a remote dataset either before or after the absorption peak to maximise the difference in anomalous signal. By collecting these data from a single crystal problems with non-isomorphism can be avoided. However, the changes resulting from anomalous signal can be very small and require very high accuracy measurements of the intensities to distinguish them so radiation damage must be avoided. Many MAD structures have been solved by replacing native methionines with selenomethionines which have a strong anomalous signal around standard MX energies (12.4 keV); but this can be detrimental to protein expression, purification, crystallisation

and gives a non-native representation of the protein which may affect interpretation of the biological function.

1.6.5. SAD Phasing

The first demonstration of single-wavelength anomalous dispersion (SAD) phasing was the structure of crambin by Hendrickson and Teeter (1981) using six native sulphur atoms. SAD experiments measure the differences in intensity between a Bijvoet pair (Eqn. 1.11) from a single crystal collected above or at an energy above an absorption edge.

$$\Delta F^{\pm} = |F_{PH}(+)| - |F_{PH}(-)| \tag{1.11}$$

As with soaked heavy atoms these differences between Bijvoet mates (equivalent to native and heavy atom intensities) are used to determine the heavy atom substructure by direct or Patterson methods. However, SAD data presents with an ambiguity in the phase angle as demonstrated in Figure 1.5. Using the heavy atom substructure the amplitude and phase can be calculated (F_H), giving an origin for *f"* and the measured intensities of the Bijvoet pairs (F^+_{PH} and F^-_{PH}). Using a Harker construction to draw circles of radius F_{PH} and F^-_{PH} at origins *f"*(±) we see two potential solutions for the phases (Fig. 1.5). The phase ambiguity can be resolved by density modification which should produce an outline of a map distinguishable from the background noise when using the correct enantiomorph.

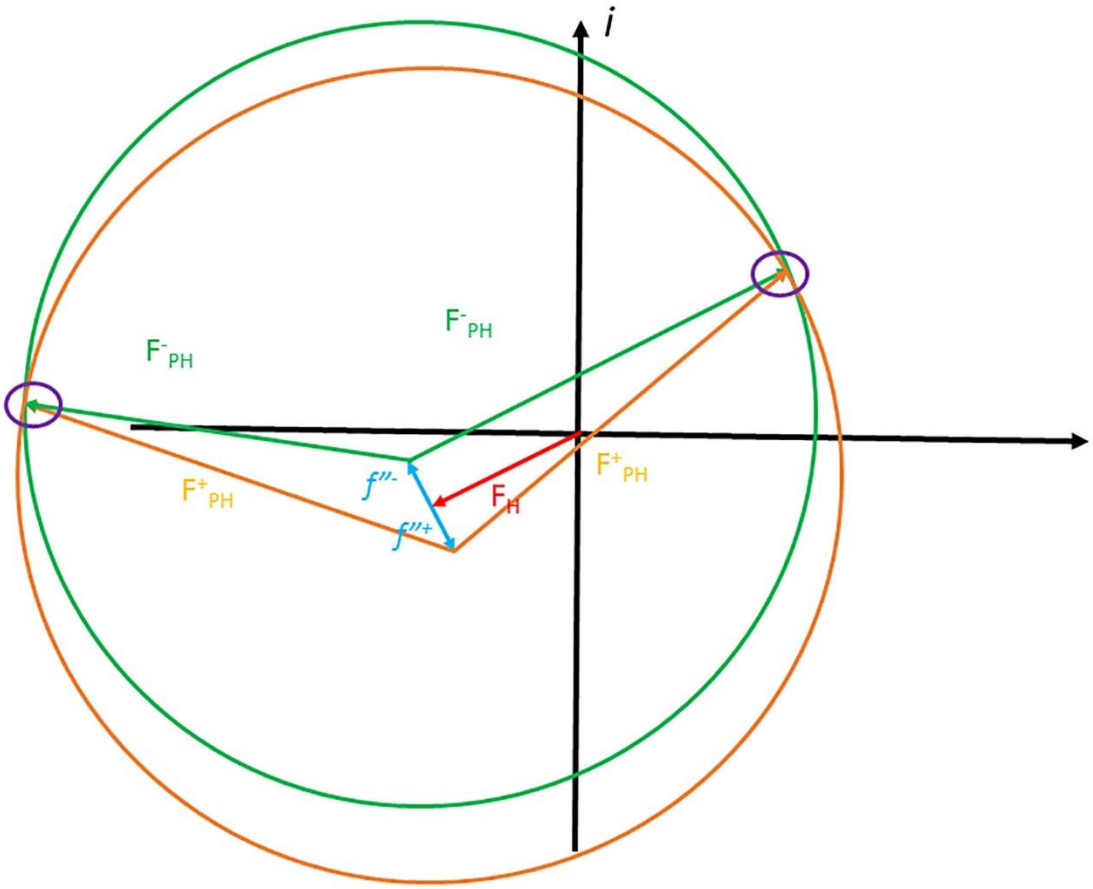


Fig. 1.5: A Harker construction demonstrating the phase ambiguity resulting from SAD phasing with the initial phase and amplitude of the anomalous substructure (F_H) in red, *f″* in blue, the measured intensities of the positive and negative Bijvoet pairs (F^+_{PH} and F^-_{PH}) and their corresponding circles in orange and green, respectively, and the two potential solutions circled in purple.

1.6.6. Long Wavelength MX for Sulphur-SAD

As discussed earlier for MAD phasing, the largest anomalous signal for any element is measured at or above its absorption edge. The absorption edge of sulphur is around 5 Å, a much longer wavelength than most standard MX beamlines can reach (0.8-2.5 Å). However, it has been shown that anomalous differences from native sulphurs can be measured using 1.54 Å X-rays from a copper rotating anode (Hendrickson and Teeter, 1981) and using synchrotron radiation between 1.54 Å (Dauter et al., 1999) and 2.5 Å (Liu et al., 2000). However, automated S-SAD phasing is still not routinely successful on standard MX beamlines (Doutch et al., 2012). An estimate of the anomalous

signal that can be obtained from native sulphurs in a protein crystal at different X-ray energies can be determined using the Bijvoet ratio,

$$BR = \left(\frac{N_A}{2N_T}\right)^{\frac{1}{2}} \left(\frac{2f''_A}{Z_{eff}}\right)$$

(1.12)

where N_A is the number of anomalous scatterers in the unit cell, N_T is the total number of atoms in the unit cell, *f"* is the anomalous scattering factor at a set energy and Z_{eff} is the normal scattering power (6.7 e- at 2θ = 0). It has been suggested that a Bijvoet ratio of at least 0.6% is necessary for successful phasing by SAD (Wang, 1985). Although use of longer wavelength (1.5-3 Å) and long wavelength (>3 Å) X-rays have the advantage of increased anomalous signal for lighter atoms such as sulphur, phosphorous, calcium and chlorine there are several downsides that must be taken into account.

As wavelength increases scattering increases proportional to λ^2 but absorption by the crystal, surrounding solvent, sample mount and air also increase proportional to λ^3 (Djinovic Carugo et al., 2005). Based on this ratio it has been argued that smaller crystals will scatter much more efficiently at longer wavelengths, providing radiation damage does not occur (Teplyakov et al., 1998). However, reducing radiation damage can prove difficult as measurement of the weak anomalous signal from native sulphur requires highly redundant data (Dauter and Adamiak, 2001, Weiss, 2001). The advent of fast Pixel array detectors (Broennimann et al., 2006b) means highly accurate data can now be collected before the onset of radiation damage (Owen et al., 2014b), as well as with minimal readout noise. Significant improvements in data processing software have also helped get the most out of the small anomalous signal measured (Kabsch, 2010b, Adams et al., 2010, Sheldrick, 2010). These developments have brought about a resurgence in S-SAD phasing with the use of multi-crystal averaging (Liu et al., 2012, El Omari et al., 2014) and multiple low dose data collections on a single crystal (Weinert et al., 2015). Although the state of S-SAD has drastically

improved over the years, standard MX beamlines are unable to use long wavelength X-rays to their full potential due to absorption and scattering of X-rays in air, absorption in the crystal and the larger diffraction angles from the X-rays (Djinovic Carugo et al., 2005). In an attempt to harness the power of long wavelength X-rays, Diamond Light Source have developed the first in vacuum beamline for long wavelength MX, I23.

1.7. I23 – In-vacuum MX

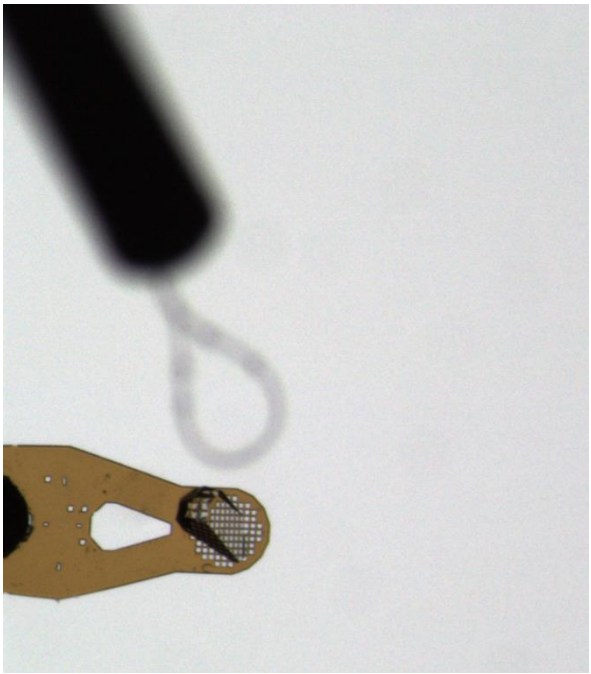
To reduce the effects of background noise on the detector from air scatter as well as the absorption of the X-ray beam in air I23 has been designed to operate in vacuum, with the entire experimental setup in a large vacuum chamber. By performing diffraction experiments in vacuum scattering of X-rays by air is eliminated, leaving only background scatter from the crystal and sample mount. The signal-to-noise ratio can also be improved by removing non-crystalline solvent surrounding the crystal which causes water rings to appear on the diffraction image. Chapters Two and Three of this thesis deal with the effects of longer wavelength X-rays in the presence and absence of excess surrounding solvent on data quality. Performing diffraction experiments in vacuum means conventional cooling of protein crystals with a stream of liquid nitrogen is not possible so I23 has adopted a conductive cooling mechanism. Extensive work has been carried out on the viability of cooling by conductance through the goniometer and specially designed sample mounts (Warren et al., 2013). Cooling by conductance also has the added benefit of minimising sample vibrations from the cryo-stream which can have a negative effect on the signal-to-noise ratio (Alkire et al., 2008, Alkire et al., 2013). Finally, with respect to improving signal-to-noise, a semi-cylindrical Pilatus 12M detector has been designed to allow rapid data collections with minimal readout noise and to account for the large scattering angles from long wavelength X-rays.

Although absorption has been limited through removal of air and surrounding solvent absorption is still a limiting factor for long wavelength MX with respect to the crystal itself. Absorption correction accounts for the attenuation of X-rays travelling through the crystal, however,

most absorption correction programs use a rough estimation of crystal volume. I23 will employ an X-ray tomography system to allow accurate measurement of the crystal volume post data collection for more accurate absorption correction. Tomographic reconstruction of macromolecular crystals has been demonstrated by Brockhauser et al. (2008) on lysozyme and DNA crystals. X-ray tomography has also been used as a low dose alternative to the grid scans for location of small membrane protein crystals cooled in opaque lipidic cubic phase mother liquor (Warren et al., 2013).

As of the writing of this thesis (July 2015) I23 has not entered user operation and only limited work has been done in house for MX. This thesis is concerned with the development of sample handling techniques to allow collection of superior quality data in the face of increased absorption and background scatter when approaching longer wavelengths for native S-SAD phasing. Chapters two and three cover the use of a humidity controlled device (HC1) (Sanchez-Weatherby et al., 2009b) and a custom built sample handling setup to remove excess surrounding solvent from macromolecular crystals for diffraction experiments at standard and longer wavelengths. Robust statistical tests were applied to assess the effects of solvent on data quality. Chapters four and five concern the expression, purification and crystallisation of novel protein targets for S-SAD phasing on I23. However, due to the technical difficulties in development of the beamline these crystals were not tested on I23 but can be reproduced for testing at a later date. Chapter six presents the structural and biochemical analysis of a novel EAL domain protein from *Pseudomonas aeruginosa* involved in the regulation of biofilm formation. EAL domain protein crystals were also used in data quality tests presented in chapter three. Finally, Chapter seven demonstrated the advantages of high quality diffraction data by investigating the reaction mechanism of *Achromonacter cycloclastese* nitrite reductase with atomic resolution crystal structures collected using serial data collection on a single protein crystal to observe conversion of nitrite to NO in the crystal.

Chapter 2 – Sample Preparation for Long Wavelength MX



2.1. Introduction

As modern macromolecular crystallography (MX) is used to tackle more and more difficult projects, there is a pressing need to improve the quality of diffraction data. Development of synchrotron instrumentation has greatly improved the reliability, accuracy and speed of data collection to the point that sample quality is becoming the more significant stumbling block. The major limiting factors affecting diffraction data include crystal quality, radiation damage and the signal-to-noise ratio. Improving crystal quality requires the systematic optimisation of crystallisation conditions and cryo-protection which can be very time consuming and may require rescreening to find alternative crystallisation conditions. Radiation damage results from interactions of the X-ray beam with the electrons in the crystal. Although X-ray diffraction is an elastic process, X-ray absorption and Compton scattering deposit energy in crystals, causing damage and a reduction in diffracting power (Teng and Moffat, 2000, Garman and Weik, 2015). These effects are wavelength dependent at typical MX energies (12.4 keV), with damage proportional to absorbed X-ray dose (Owen et al., 2006a, Murray et al., 2005, Henderson, 1990). The development of X-FEL technology allows a “collect-before-destroy” approach to overcome the effects of primary radiation damage caused by direct interaction of X-ray radiation with electrons to break chemical bonds (Neutze et al., 2000, Teng and Moffat, 2000). However, access to X-FELs is very limited so does not present a feasible alternative to synchrotron radiation for the MX community. Secondary damage results from chemical reactions of highly reactive species produced by ionising radiation, such as hydroxyl radicals, hydrated electrons and H atoms (Burmeister, 2000, Teng and Moffat, 2000, Garman, 2003a, Murray et al., 2004, Owen et al., 2006a). These effects can be reduced by slowing the thermal diffusion of radicals through

data collection performed at cryo temperatures (Hope et al., 1989) or by outrunning damage with fast detectors and very short exposure times (Owen et al., 2014a).

Maximising the signal-to-noise ratio has been addressed through improved data collection techniques, largely enabled by modern synchrotron instrumentation such as pixel array detectors (Broennimann et al., 2006a) and microfocus beams (Evans et al., 2007). The combination of the fine ϕ-slicing method (Mueller et al., 2012) and precise targeting with microbeams (Sanishvili et al., 2008) have both contributed to a significant reduction in background scatter from surrounding solvent and sample mounts. In addition to exploiting these advances in synchrotron technology, careful sample preparation can go a long way towards improving the signal-to-noise ratio during diffraction experiments.

The principal causes of background scatter are the excess solvent around frozen crystals and the sample mounts. Although this background scatter is not usually considered a critical issue in standard MX experiments, for the long wavelength experiments required for accurate phasing and in microcrystal MX, where signals are significantly weaker, background scatter becomes a limiting factor. As a result, in an effort to improve the signal to noise ratio, a variety of sample handling techniques have been developed as alternatives to traditional nylon and Kapton mounts.

The “capillary-top mounting” or “loopless method” was developed to exploit the use of Cr Kα radiation (2.29 Å) and reduced background scatter to measure the weak signals associated with S-SAD phasing (Kitago et al., 2005b, Watanabe et al., 2006, Kitago et al., 2010b). Capillary mounts with removable nylon loops (Fig. 2.1c-d) were incorporated into a semi-automated setup, which removes surrounding solvent by aspiration and flash freezes protein crystals in the cryostream in one motion (Fig 2.1a and b) (Kitago et al., 2010b). This

method has seen some success in structure solution by S-SAD and presents a good case for phasing by in-house Cr Kα radiation (Watanabe et al., 2006).

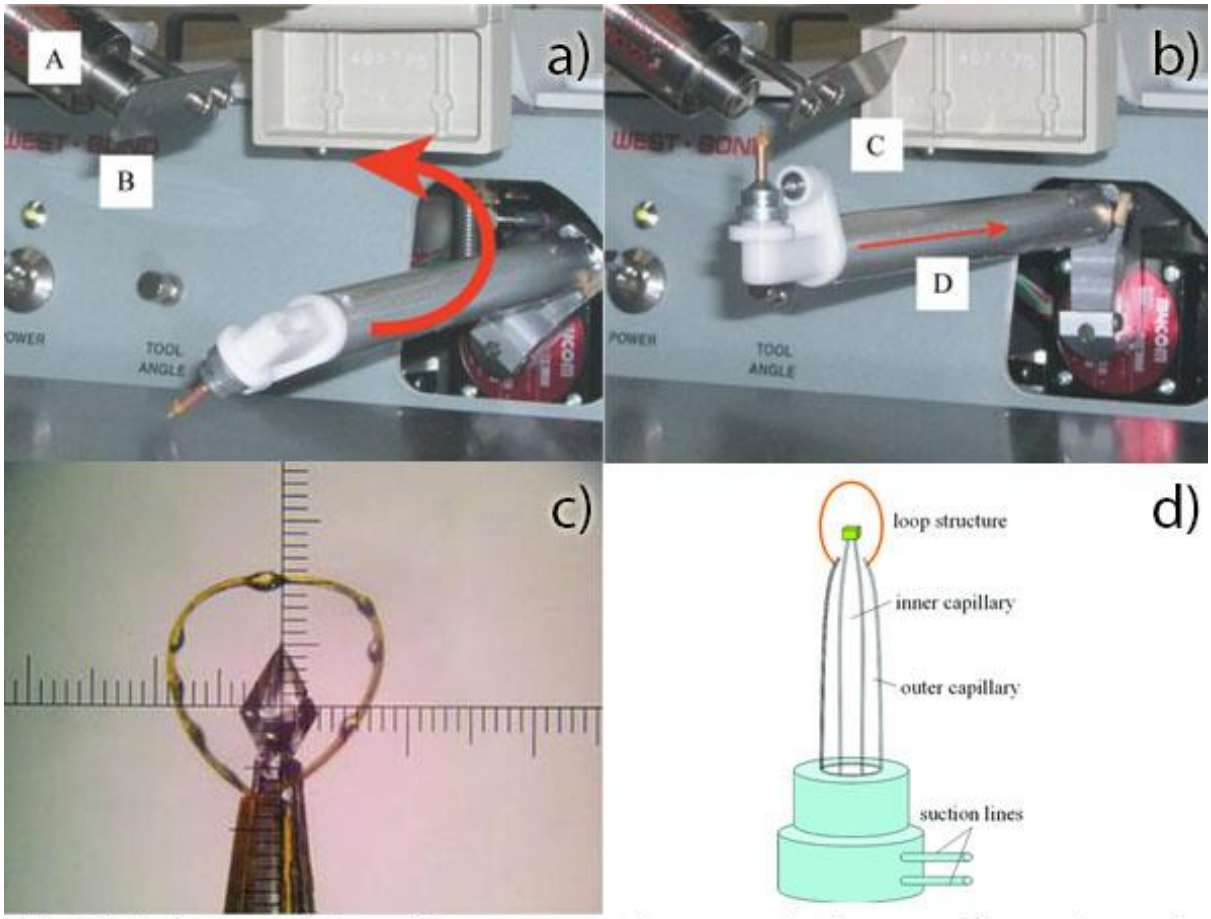


Fig 2.1: A close up of the capillary-top mounting system in the crystal harvesting position (a) with the cryo-stream (A) and cryo-stream shutter (B) in the closed position; and in the crystal freezing position (b) with the cryo-stream shutter open (C). The direction of the suction line (D) for aspiration is indicated by the red arrow in (b). An example of a frozen thaumatin crystal mounted using the capillary-top mounting method is shown in (c), each marker represents 50 µM. (d) A conceptual design for a dual capillary mount for handling smaller crystals. Figure adapted from Kitago et al. (2010d).

Another technique designed to improve the signal-to-noise ratio is the Crystal Catcher, designed by Kitatani et al. (2008b) and further tested by Mazzorana et al. (2014). The Crystal Catcher allows loop free mounting of crystals on an adhesive ejected from a quasi-SPINE standard pin using a mechanical pencil type mechanism (Fig. 2.2) (Kitatani et al., 2008a, Kitatani et al., 2008b, Kitatani et al., 2009).

Wierman et al. (2013) have also developed a technique using atomically thin graphene sheets to coat protein crystals. Crystals are wrapped by pipetting crystals into a drop with a graphene sheet floating at the water air interface of the drop (Fig. 2.3a). As the crystal floats to the bottom of the drop a nylon loop is used to position the crystal in the graphene sheet which wraps around the crystal as it is pulled from the drop (Fig. 2.3b) (Wierman et al., 2013).

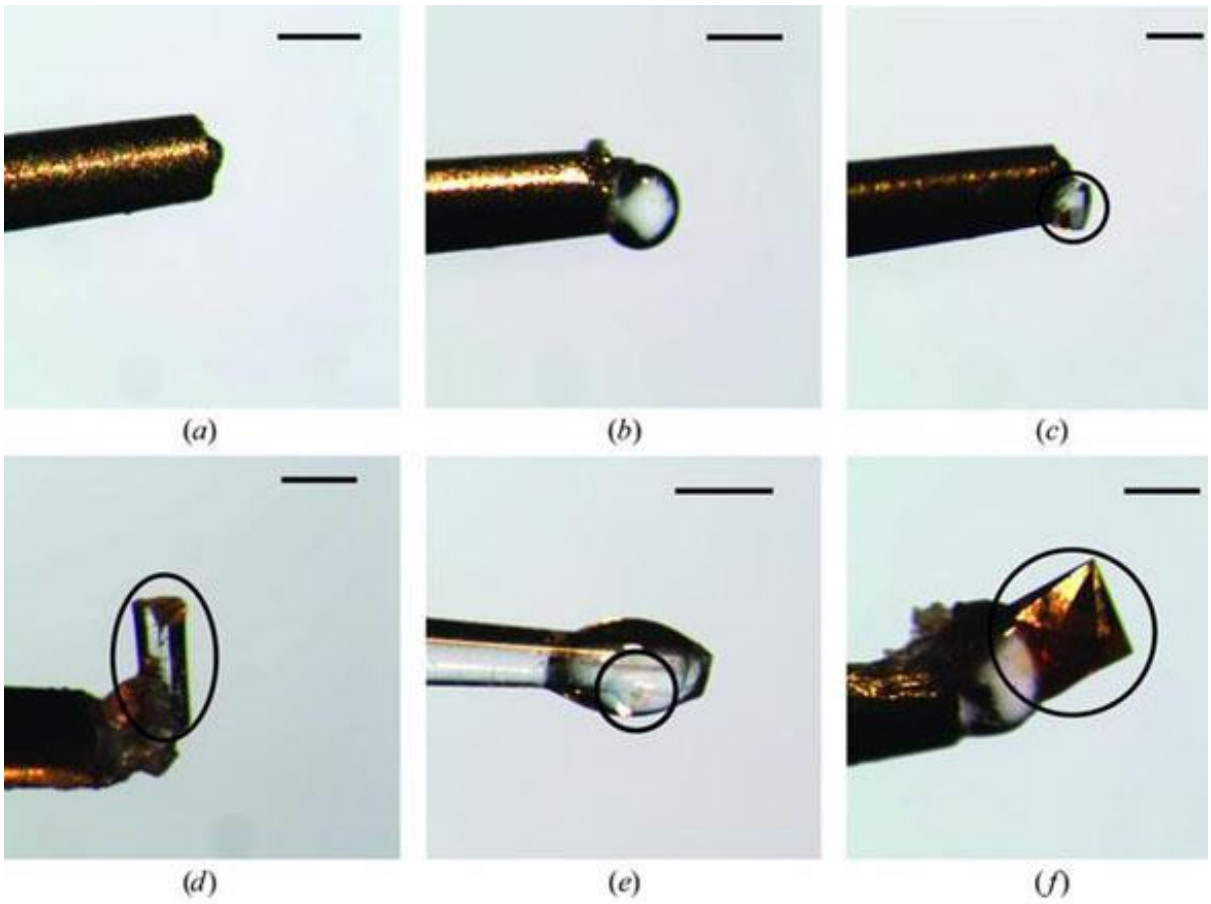


Fig. 2.2: Examples of crystal harvesting with the "Type A" adhesive on metal pins with a small amount of adhesive (a), a large amount of adhesive (b), a small lysozyme crystal (c), a large thermolysin crystal (d), a trypsin crystal mounted with a CT-400 glass pin (e) and a holoferritin crystal (f). Black circles indicate a crystal and black bars measure 200 µm. Figure adapted from Mazzorana et al. (2014).

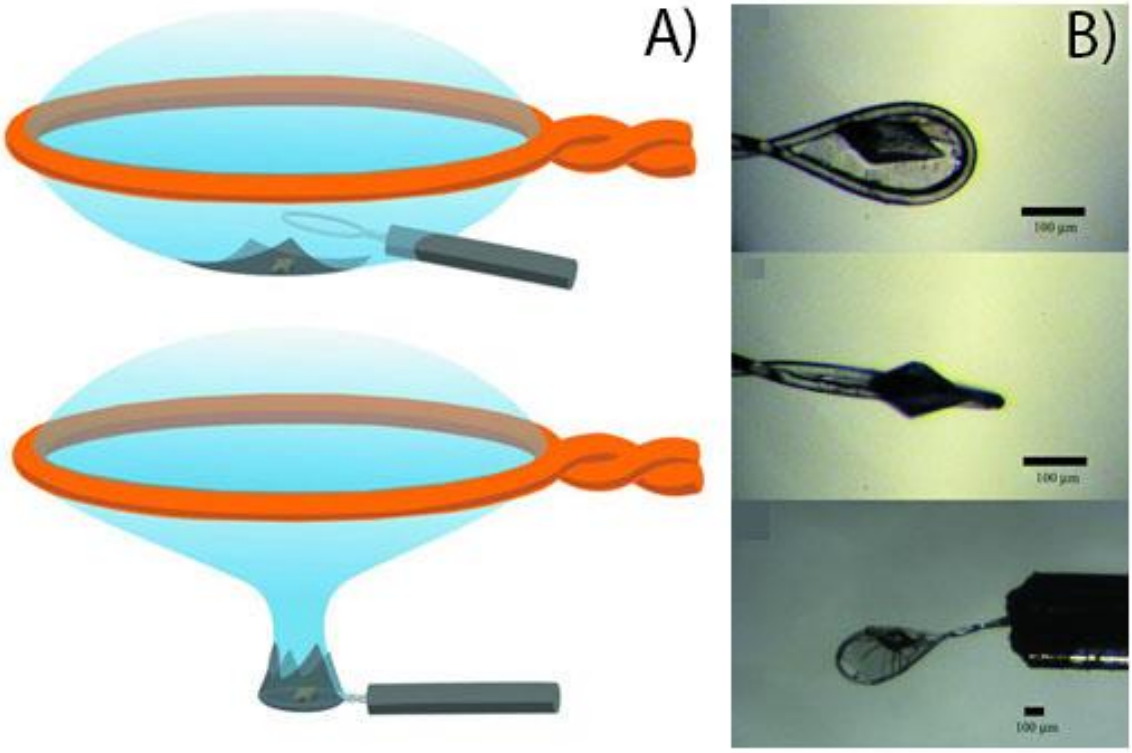


Fig. 2.3: A graphic depicting the graphene mounting process (A) where a nylon loop is passed through a crystal floated on top of a layer of graphene; and a mounted thaumatin crystal (B). Figure adapted from Wierman et al. (2013).

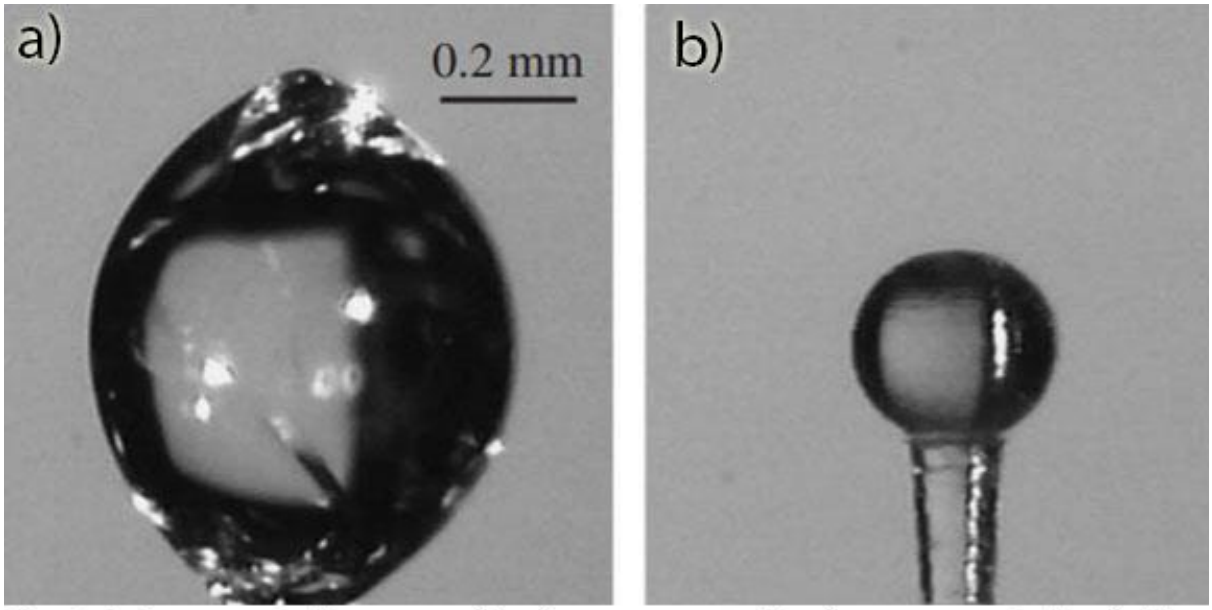


Fig. 2.4: A cryo-cooled hen egg white lysozyme crystal in a loop surrounded by buffer before (a) and after being processed by PULSA to produce a 0.3 mm diameter protein crystal ball (b). Figure adapted from Kitano et al. (2005a).

The use of lasers has also been explored for sample preparation in an effort to increase the signal to noise ratio. Pulsed UV laser soft ablation (PULSA) has been used on crystals of soluble protein to cut and shape crystals, minimising surrounding solvent without damaging the structure of the crystal (Kitano et al., 2004, Kitano et al., 2005a, Watanabe et al., 2006). The technique has been extended to processing of membrane proteins and the removal of nylon loops using PULSA (Kitano et al., 2005b). CrystalDirect (Cipriani et al., 2012) also employs laser-induced photoablation for sample preparation with a femtosecond laser used to excise protein crystals grown on ultrathin films. This technique aims to completely automate crystal harvesting and reduce background scatter in the process (Fig. 2.5) (Cipriani et al., 2012). The work by Cipriani et al. (2012) corroborates the findings that the use of PULSA has no detrimental effects of crystal quality.

Although not originally designed for the purpose of reducing background scatter to improve the signal-to-noise ratio, both the free mounting system (FMS) (Kiefersauer et al., 1996, Kiefersauer et al., 2000) and humidity controlled device (HC1) (Sanchez-Weatherby et al., 2009a) allow removal of excess surrounding solvent while maintaining protein crystals in a humid air stream. Both were designed to exploit the dynamic nature of protein crystals by using dehydration to alter crystal packing and improve the diffraction power of crystals; with an ever growing list of success stories (Russo Krauss et al., 2012, Deng et al., 2012, Newman, 2006, Bowler et al., 2006b, Heras et al., 2003, Heras and Martin, 2005, Kalinin et al., 2005). The FMS removes a crystal from a nylon loop using a micropipette; the crystal is maintained within a stream of humid air to allow online dehydration experiments (Fig. 2.6). The relative humidity of the air stream is regulated by mixing two air streams of 0 and 100% relative humidity (Kiefersauer et al., 2000). Solvent is removed by aspiration in a similar

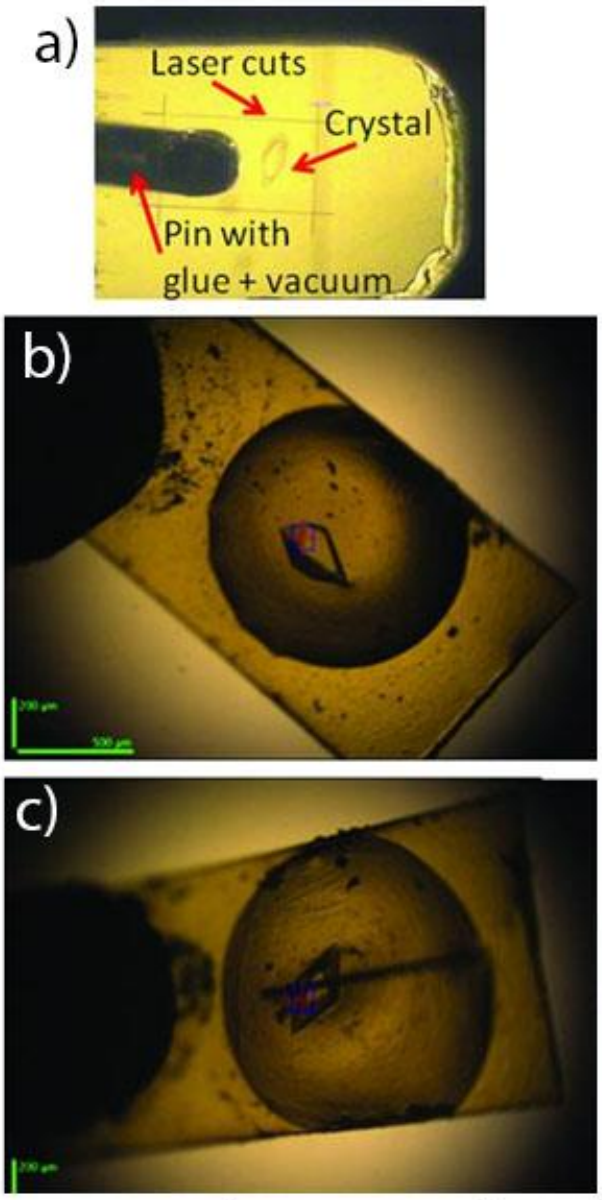


Fig. 2.5: CrystalDirect automated crystal harvesting system. Crystals are grown on a crystallisation film and excised using a laser and specialised pin with glue and a vacuum (a-b). Crystals can be shape further with the laser to reduce surrounding mother liquor (c).

fashion to the capillary top mounting method. Although the FMS has been successful in many cases, it has proven difficult to routinely or easily incorporate its bulk into the crowded synchrotron beamline environment.

This has led to the development of a more suitable online device, the HC1, which uses a nozzle based on standard cryostream technology for delivery of the humid air stream (Fig. 2.7) (Sanchez-Weatherby et al., 2009a, Russi et al., 2011). Protein crystals can be harvested on standard sample mounts and placed in the humid air stream, where excess solvent can be removed to allow more efficient dehydration (Sanchez-Weatherby et al., 2009a). The HC1 regulates humidity in a similar manner to the FMS, using calculation of the dew point to set the final relative humidity (Kiefersauer et al., 2000). However, differences arise between the two systems in how the dew point is achieved. Instead of mixing two air streams, the HC1 use a condenser paired with a Peltier device (Fig.

2.7b) to cool the incoming air to the calculated temperature and remove excess water though the condenser (Sanchez-Weatherby et al., 2009a). In this study the HC1 has been used to maintain protein crystals, as done previously (Sjogren et al., 2002, Sjogren and Hajdu, 2001b, Sjogren and Hajdu, 2001a), to allow removal of excess surrounding solvent, to assess the effects of manual handling and frozen solvent on data quality.

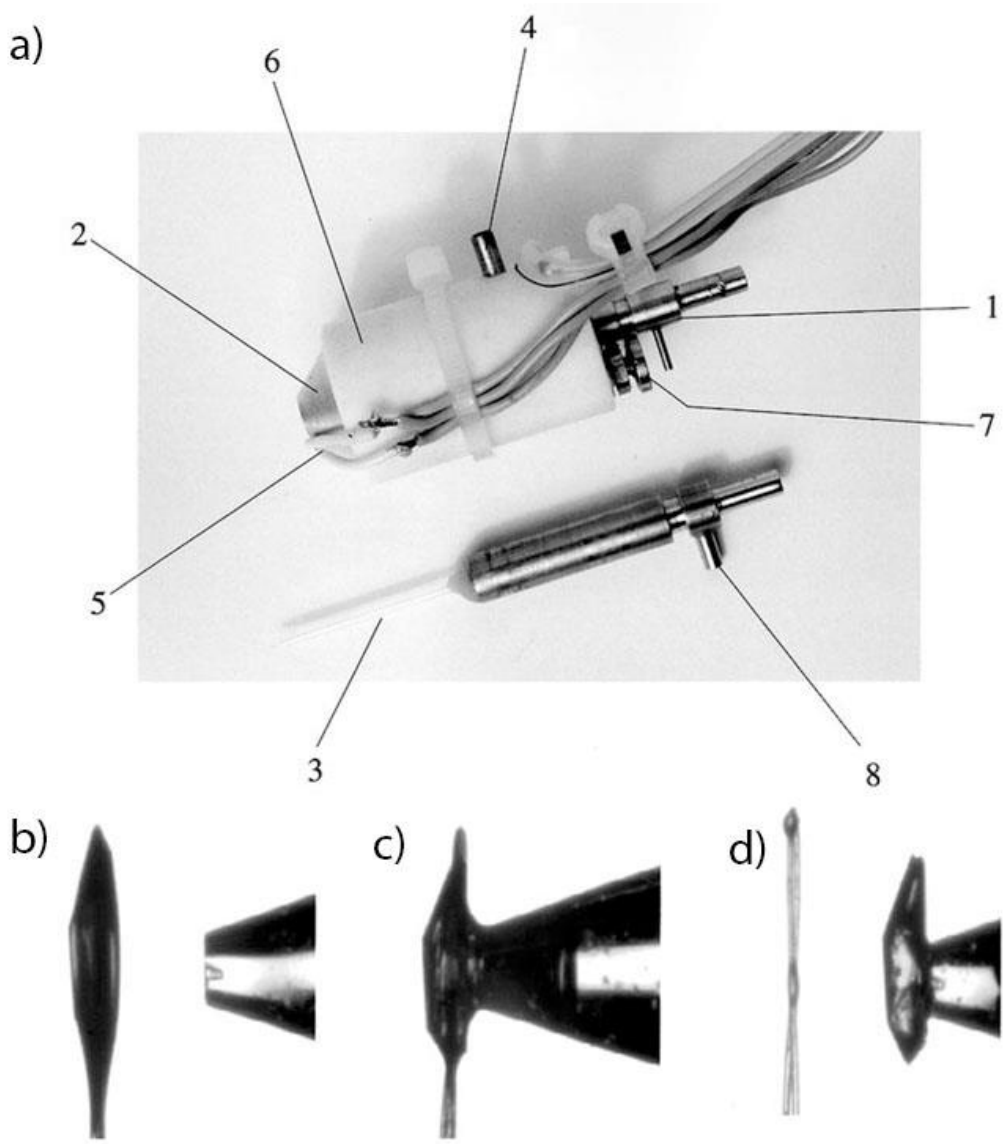


Fig. 2.6: The Free-mounting system (FMS). The main components of the FMS are the insert (a1) and head part (a2) which holds the micropipette (a3) in place. Crystals are sucked onto the micropipette from a nylon loop (b-d) by the vacuum line (a8) and maintained using a humid air stream fed into the head part (a4). The temperature is measured with a Pt100 ceramic sensor (a5) and the outer head part is shielded with Teflon for better temperature control. Micropipette position is set with a screw (a7) (Kiefersauer et al., 2000).

All of the studies mentioned above have presented findings based on comparison of a very small number of (or even just single) protein crystals between treatments, without validation by an appropriate statistical test (Kitago et al., 2010a, Kitago et al., 2005b, Kitatani et al., 2008b, Mazzorana et al., 2014, Wierman et al., 2013, Cipriani et al., 2012). However, studying and accurately evaluating relatively small effects on data quality is very difficult to achieve due to radiation damage and the intrinsic variability of crystals, even if grown and harvested under the same conditions and in the same drop (Nowak et al., 2009).

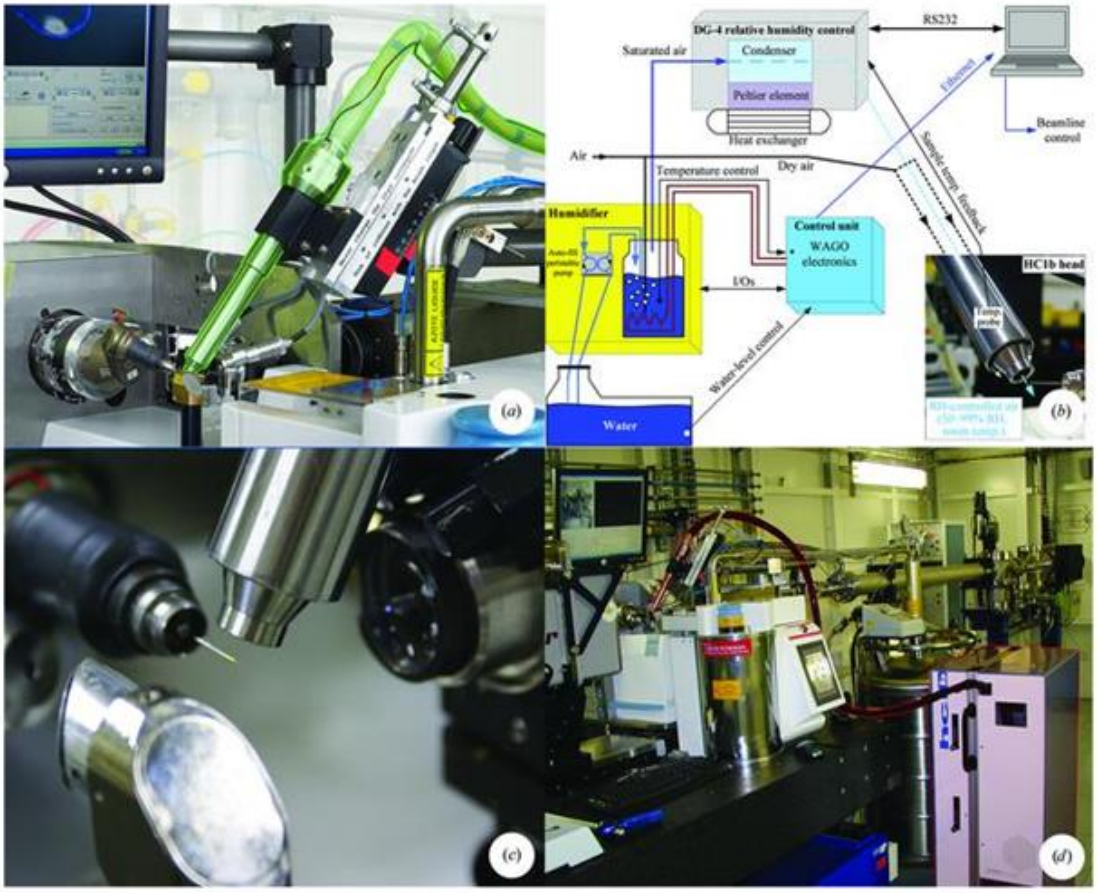


Fig. 2.7: The HC1b humidity-control device. The HC1b nozzle, in green, mounted on an ESRF MX beamline (a) with an enlarged view of the sample environment (c). (b) A schematic of the HC1b with water reserve leading into a humidifier (yellow); saturated air then passes into the DG-4 relative humidity control box (grey) where the condenser (cyan) and Peltier element (purple) are housed. Humid air at the desired relative humidity is then directed with the nozzle to the sample environment. (d) The HC1b installed in an experimental hutch. Figure adapted from Sanchez-Weatherby et al. (2009).

In this study, a systematic statistical approach was used to determine whether manual handling and manipulation of crystals in a humidity-controlled stream can be performed at no cost to the integrity of the sample. To prove this concept, large populations of macromolecular crystals were transferred between conventional sample mounts, using micromanipulators to allow precise movement, and the HC1 to preserve the integrity of macromolecular crystals and enable removal of excess surrounding solvent. crystal transfer was facilitated by the removal of excess surrounding solvent with a fine paper wick (Warkentin and Thorne, 2009), giving the added benefits of reduced background scatter and improved cooling rates during cryo-protection. To assess the effects of manual handling and the presence of excess solvent, two statistical tests were applied to the crystal/data populations: Student's t-test (Student, 1908) and the Mann-Witney-U-test (Mann and Whitney, 1947). Both tests examine the null hypothesis that the population of transferred crystals and the population of crystals fished and mounted by conventional methods show no significant differences in data quality. Initially, experiments during data collection were conducted at λ = 1 Å, a typical wavelength used in MX, to ensure that sample handling, exposure to the humid air stream and wicking of excess solvent can be performed at standard conditions without damaging macromolecular crystals. These experiments and tests were a precursor and background to testing the effects of excess solvent using long wavelength data collection (see Chapter 3). At λ = 1 Å it is expected that absorption will be negligible, so any detrimental effects on data quality will be the result of manual handling and/or exposure to the humid air stream.

2.2. Methods

2.2.1. Protein Crystallisation

Hen egg white lysozyme (HEWL), *porcine* insulin, *Thaumatococcus daniellii* thaumatin and horse-spleen ferritin were selected to study the effects of manipulating protein crystals and removing excess solvent in a humidity controlled environment. These proteins are all commercially available, have well defined crystallisation protocols and are grown in optimised cryo-protection conditions, thus eliminating the additional manual step of determining cryo-protection conditions, which could introduce random errors into this systematic study. All proteins were purchased from Sigma-Aldritch (St Louis, Missouri, USA) and underwent no further purification. Crystals were grown in sitting drops by the vapour diffusion method at 19°C over several days (Fig. 2.8).

Tetragonal HEWL crystals were grown in 0.8-1.2 M NaCl, 50 mM sodium acetate pH 4.6 and 25% ethylene glycol (v/v) following the protocol by Murray and Garman (2002). 30% PEG 5K (w/v) was substituted for 25% (v/v) ethylene glycol for cryo-protection. *Porcine* insulin was suspended in 50 mM sodium phosphate with 10 mM EDTA pH 10.2 and equilibrated against 30% ethylene glycol and 50 mM sodium phosphate with 10 mM EDTA pH 10.2 in a 1:1 ratio to produce cubic crystals, adapted from the batch method by Meents et al. (2007). Tetragonal thaumatin crystals were grown by equilibrating 80 mg/ml thaumatin in 100 mM phosphate buffer pH 6.5 against 30% (w/v) Na/K tartrate, 100 mM phosphate buffer pH 7 and 20% glycerol, adapted from the modified gel acupuncture method (Lopez-Jaramillo et al., 2001). Cubic ferritin crystals were grown in 0.1 M NaCl, 0.8

M ammonium sulphate, 10 mM cadmium chloride and 25% glycerol (v/v) as detailed by

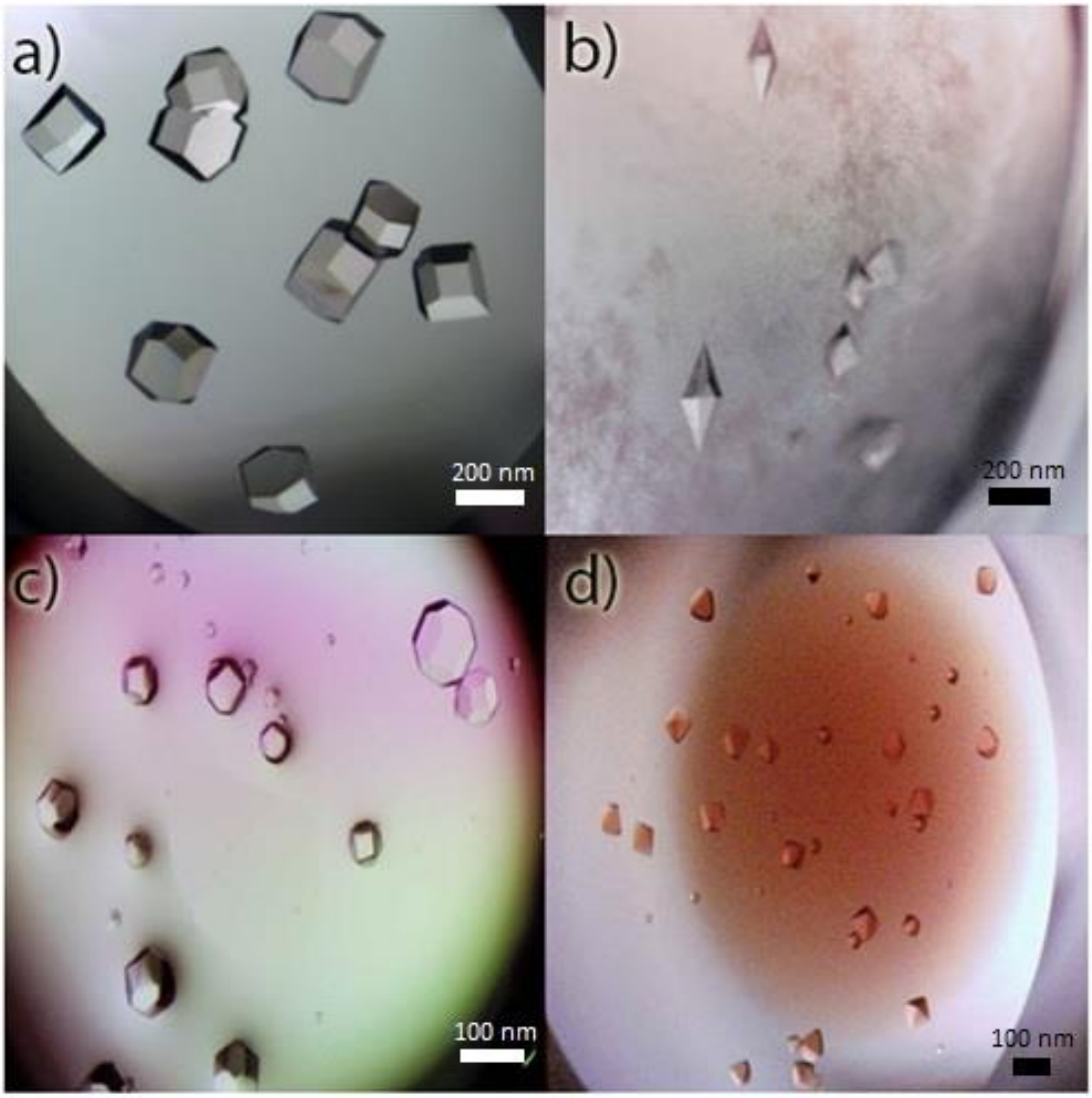


Fig. 2.8: Protein crystals for HC experiments. Tetragonal lysozyme (a), tetragonal thaumatin (b), cubic insulin (c) and cubic holoferritin (d). Average crystal sizes can be found in Tables 2.3-5.

Murray and Garman (2002).

2.2.2. The Humidity Control Device (HC1)

The main purpose of humidity control devices to date has been to dehydrate/rehydrate protein crystals to subtly alter crystal packing and ultimately improve

diffraction quality (Sanchez-Weatherby et al., 2009a, Bowler et al., 2006c, Russi et al., 2011). In my case the HC1 (Arinax, ZI de Centr'Alp, France) was used to (i) maintain protein crystals in as near a native environment as possible for the physical transfer from one mount to another and (ii) allow controlled removal of excess surrounding solvent.

2.2.3. Defining Relative Humidity

To prevent protein crystals dissolving, drying out or losing cryo protection when introduced to the humid air stream their relative humidity must be measured. Relative humidity was measured by loading a drop of reservoir solution on an empty loop in the humid air stream and monitoring the change in drop size over time. Too high a relative humidity causes the drop to grow while too low will cause the drop to shrink. The relative humidity was adjusted accordingly and the process repeated with a new drop of reservoir solution until equilibrium was reached and drop size stabilises (Sanchez-Weatherby et al., 2009a). A fresh drop was used after each humidity adjustment as the concentration of buffer components may have changed with loss or gain of water from the humid air stream. All crystal transfer experiments, described later, were carried out under these equilibrium conditions. The relative humidities of several common crystallisation compounds have been measured (Table 2.1) (Wheeler et al., 2012) and these can provide initial relative humidity targets. However, reservoir solutions with a combination of buffers, salts and PEGs, etc, were observed to vary considerably from these measured 'pure' values (Table 2.1) and thorough experimentation was therefore required.

Table 2.1: Relative humidity measurements of individual chemicals and crystallisation buffers. All protein crystals reqired no further cryo-protection. For crystals where two populations were produced (lysozyme and ferritin) measured relative humidities are presented as population 1/ population 2.

Individual buffers	Relative Humidity (%)	Crystal Buffers	Relative Humidity (%)
0.1 M NaCl	100.00	**Ferritin:** 0.1 M NaCl, 0.8 M ammonium sulphate, 10 mM cadmium chloride and 25% glycerol (v/v)	89.50/90.00
0.8 M AmSO4	99.25		
25% Glycerol (v/v)	95.00		
1.0 M NaCl	99.00	**Lysozyme:** 0.8-1.2 M NaCl, 50 mM sodium acetate pH 4.6 and 25% ethylene glycol (v/v)	89.00/89.00
25% Ethylene Glycol (v/v)	92.00		
20% Glycerol (v/v)	97.00	**Thaumatin:** 30% Na/K tartrate (w/v), 100 mM phosphate buffer pH 7 and 20% glycerol	86.00
30% Ethylene Glycol (v/v)	89.50	**Insulin:** 30% ethylene glycol and 50 mM sodium phosphate with 10 mM EDTA pH 10.2	91.00

2.2.4. Humidity Control Crystal Transfer Method

In order to genly transfer crystals from one mount to another while removing excess solvent I constructed and tested the crystal transfer setup (Fig. 2.9), which consists of two micromanipulator stages, a stationary stage, a light microscope and the HC1. Macromolecular crystals are transferred from a donor sample mount (Fig. 2.10a) on the stationary stage (Fig. 2.10b) to an acceptor sample mount loaded with reservoir solution (Fig. 2.10c) on micromanipulator 1 (Fig. 2.10d). The process is viewed from above using the light microscope (Fig 2.10e). Both donor and acceptor sample mounts are placed

approximately 1cm from the HC1 nozzle (Fig. 2.10f), within the laminar humid air stream. A fine paper wick (Fig. 2.10g) attached to micromanipulator 2 (Fig. 2.10h) is used to gently remove excess solvent.

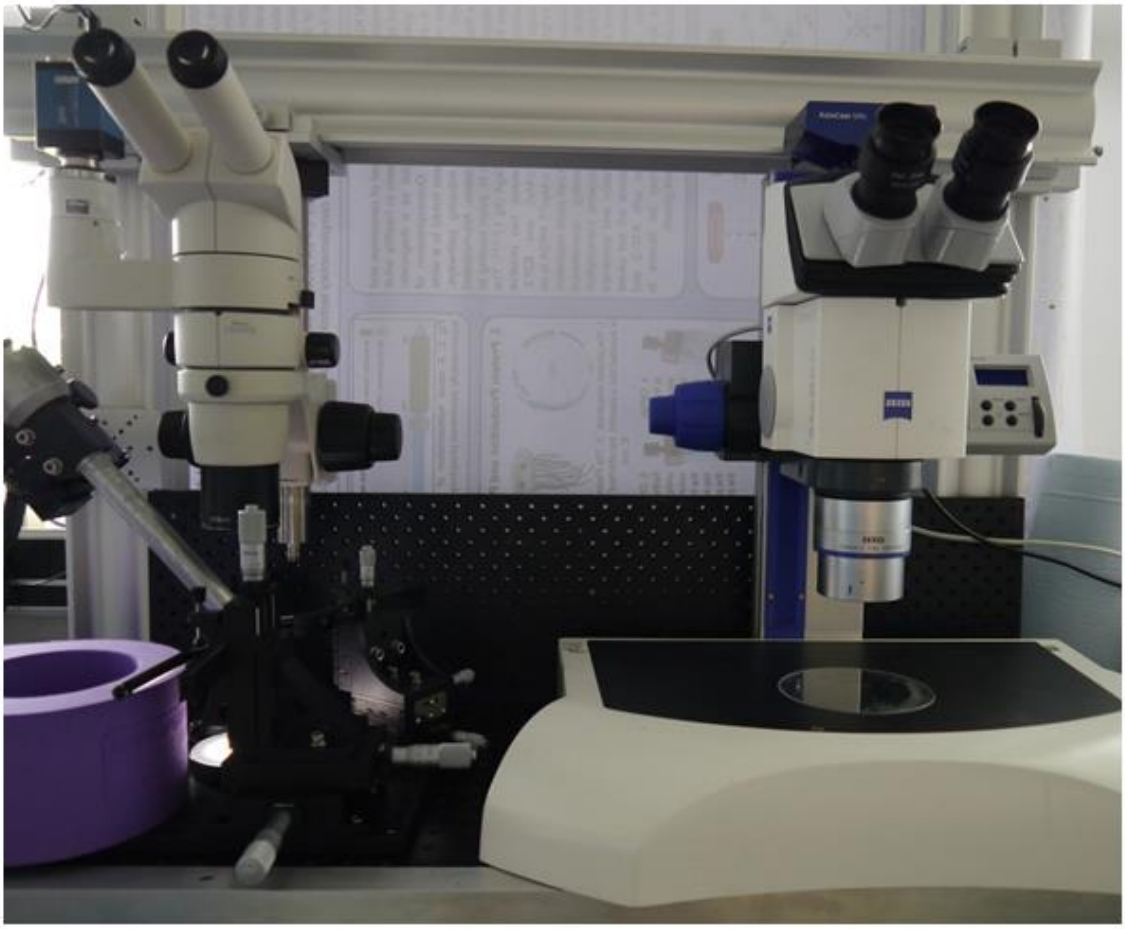


Fig. 2.9: The Crystal transfer apparatus including fishing microscope (right) and Crystal transfer set-up with HC1 (left).

The step by step transfer is depicted in Figure 2.12. A crystal is harvested from its drop using the donor mount and placed in the humid air stream opposite the acceptor mount loaded with reservoir solution (Fig. 2.12a). The acceptor mount is brought into contact with the donor mount using micromanipulator 1 (Fig. 2.12b) and the excess solvent removed by wicking (Fig. 2.12c-d). Removing the solvent from the donor mount transfers the crystal to the acceptor mount leaving minimal surrounding solvent (Fig. 2.12e-f). The

acceptor mount is then removed from micromanipulator 1 and the crystal flash frozen in the adjacent dewar of liquid nitrogen (Fig. 2.10i).

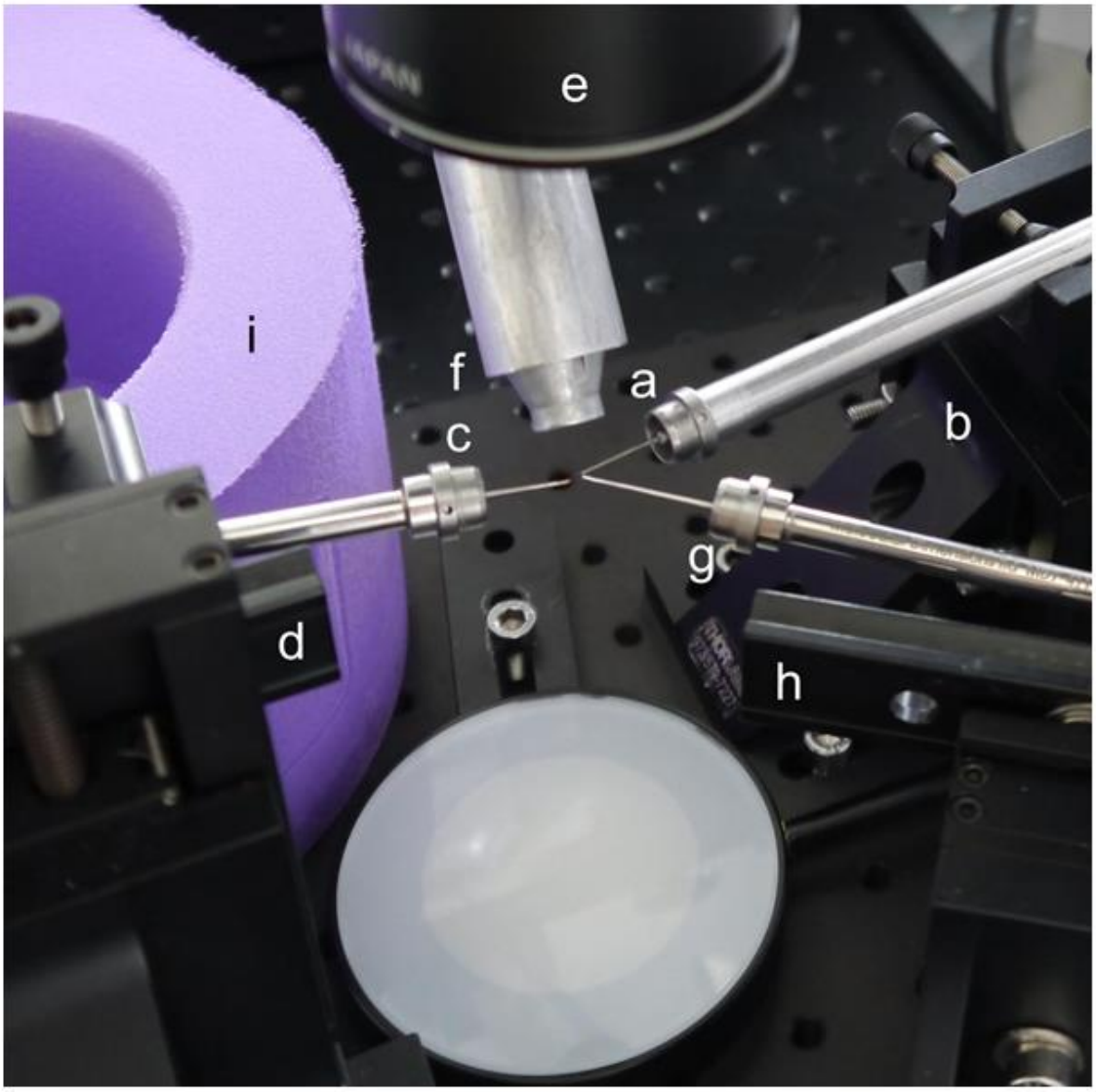


Fig. 2.10: Crystal transfer experimental setup consisting of the donor mount (a), stationary stage (b), acceptor mount (c), micromanipulator 1 (d), light microscope (e), HC1c nozzle (f), fine paper wick (g), micromanipulator 2 (h) and adjacent liquid nitrogen dewar (i).

Control crystals were harvested directly from their drops using MicroMounts or LithoLoops and immediately flash frozen in liquid nitrogen. The crystals harvested for control and HC-transfer were of comparable size for each protein crystal type (Tables 2.3-5). Crystal sizes were measured from images taken with the beamline centring microscopes.

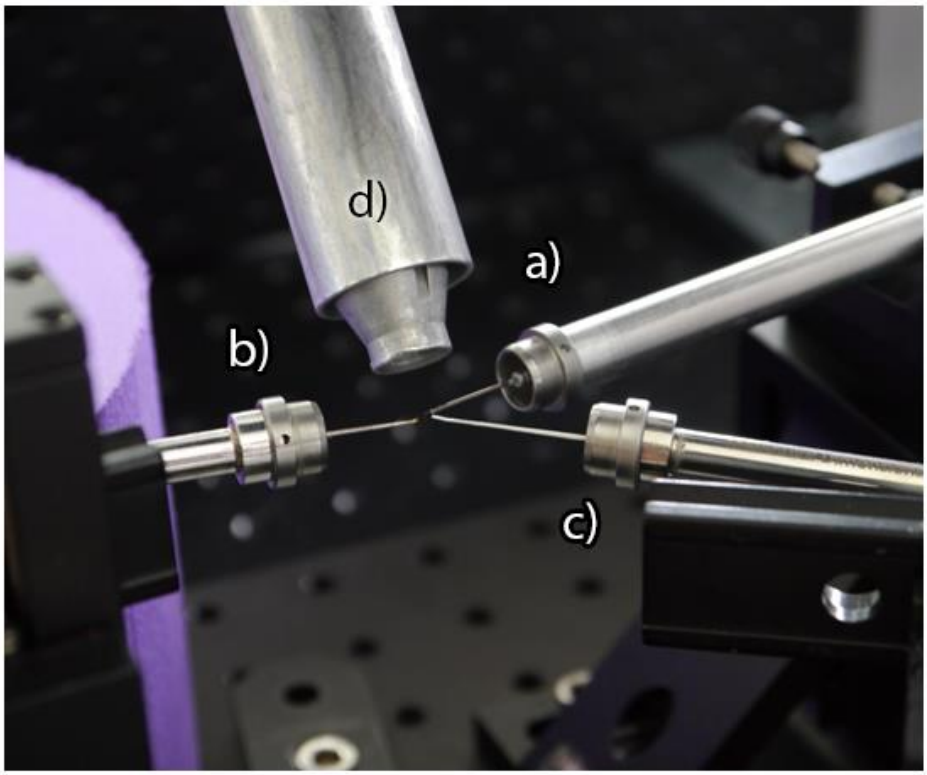


Fig. 2.11: Close up view of the donor mount (a), acceptor mount (b), paper wick (c) and HC1c nozzle (d).

2.2.5. Crystallographic Data Collection and Processing

Crystallographic data were collected over 6 visits on Diamond Light Source beamlines I03 (lysozyme 1[1], insulin and ferritin 1[1]) and I02 (lysozyme 2, thaumatin and ferritin 2). The total number of individual datasets obtained during all visits for [control/HC-transferred] crystals were as follows; lysozyme 1 [8/12], lysozyme 2 [18/16], insulin [16/13], thaumatin [7/14], ferritin 1 [16/19] and ferritin 2 [10/14]. Thus a total of 75 control and 88 HC-transferred crystals were measured on four proteins. All datasets were collected in a

[1] Lysozyme and ferritin populations were collected on two separate occasions and are named lysozyme 1 (I03), lysozyme 2 (I02), ferritin 1 (I03) and ferritin 2 (I02). Populations are not compared between primary and secondary populations (i.e. comparing lysozyme 1 with lysozyme 2) as crystals were harvested from different batches and data collection parameters varied between beamlines (Table 2.2).

single sweep to the total angular rotation stated in table 2.2, which summarises the data collection parameters.

Table 2.2: Data collection parameters.

Test Crystal	Lysozyme 1	Lysozyme 2	Thaumatin	Insulin	Ferritin 1	Ferritin 2
Beamline	I03	I02	I02	I03	I03	I02
Number of crystals	20	34	21	29	35	24
Wavelength (Å)	0.9763	0.8500	0.9795	0.9763	0.9763	0.9795
Beam Size (µm)	80 x 20	100 x 25	100 x 25	80 x 20	80 x 20	100 x 25
Detector Distance (mm)	175	147	189	177	177	230
Max. Resolution* (Å)	1.15	0.96	1.20	1.16	1.16	1.34
Total Angular Range (°)	120	120	120	90	90	90
Space Group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$I2_13$	$F432$	$F432$
Transmission (%)	20	20	20	20	20	20

All crystals were collected using a Pilatus 6M detector with a 0.1° step and 0.1 s exposure times.

*Maximum resolution as defined by the detector edge at wavelengths and detector distances stated above.

All data were indexed, integrated and scaled using the Xia2 3dii pipeline (Winter et al., 2013), XDS_ASCII.HKL files were then re-scaled with XSCALE (Kabsch, 2010a, Kabsch, 2010b) and merged with AIMLESS (Evans and Murshudov, 2013). Maximum high resolution limits were determined based on mean $I/\sigma(I) > 2$ in the highest resolution bin. All data were then processed through a second round of AIMLESS using a common high resolution cut-off specific to each protein crystal type and visit (Table 2.3-5). The lowest average maximum high resolution limit of the control and HC-transferred populations was used as the common high resolution limit for final processing and statistical analysis of each

protein type. All processing was performed using simple bash scripts to process all datasets in one sweep using the Diamond Light Source cluster (Appendix: Script Examples). Both lysozyme 1 and lysozyme 2 crystal poopulations diffracted beyond the resolution limit set by detector distance and wavelength so were processed to the maximum resolution limit available with data over 80% complete in the highest resolution bin. In both visits, ferritin data processed to the average maximum high resolution limit presented with high R_{merge} (>150%) and R_{pim} (>50%) values, despite having an average $I/\sigma(I) > 2$, so the high resolution limit in these cases was reduced to give an average R_{merge} around 100%, as used by Xia2 (Winter et al., 2013). Merging statistics for both control and HC-transferred crystals for all protein types were analysed to provide averages and standard deviations (Table 2.3-5) for a formal statistical treatment of the data.

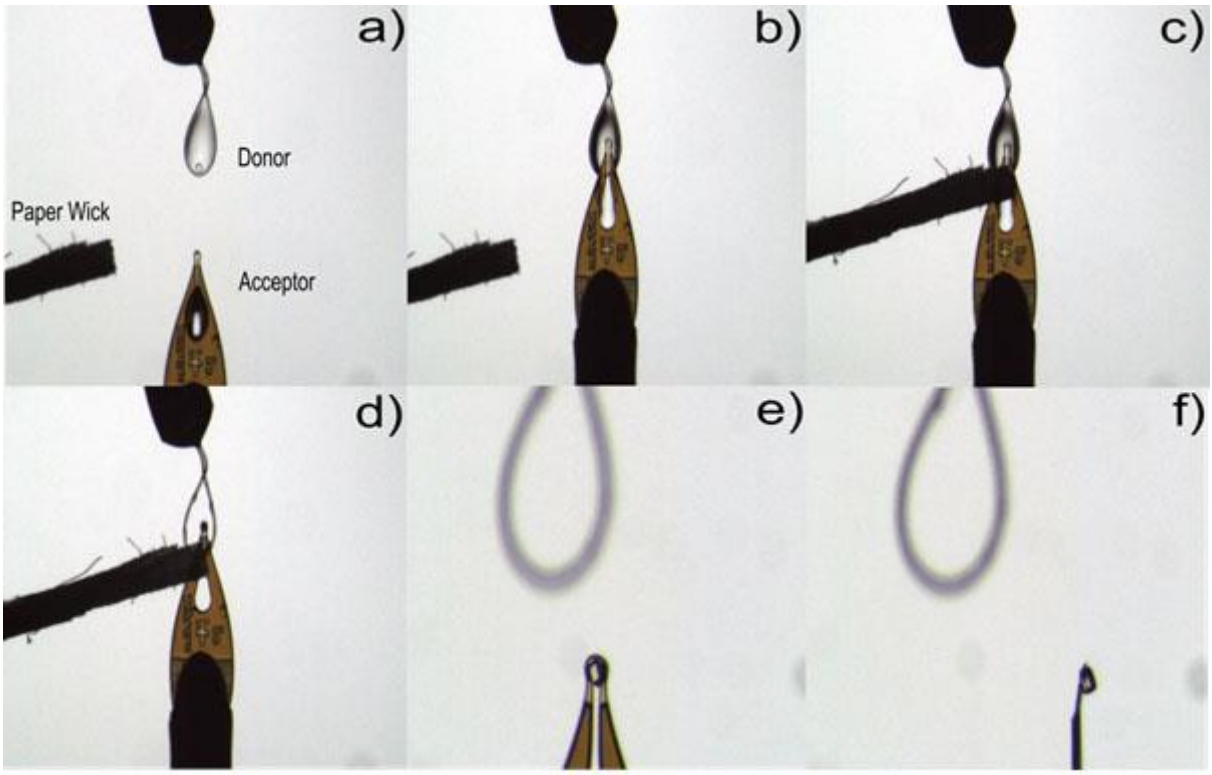


Fig. 2.12: A microscope view of the step wise transfer of a 50 µm insulin crystal from a nylon CryoLoop to a Kapton MicroMount by removing excess solvent. Donor and acceptor mounts were placed on wands pointing into the humidity stream (a) and brought into contact using micromanipulators (b). Excess solvent was then wicked away using a fine liquid wick (c and d), transferring the insulin crystal to the acceptor mount with significantly reduced surrounding solvent (e and f).

2.2.6. Statistical Analysis

Owing to the effects of radiation damage and the intrinsic variability of most macromolecular crystals it is not possible to study with any certainty the small effect that crystal handling or manipulation has on diffraction data quality using only one or a small number of samples. Historically a significant number of method development studies have drawn sometimes far reaching conclusions based on single crystal measurements or from small populations of X-ray diffraction datasets (Kitago et al., 2010a, Kitago et al., 2005b, Kitatani et al., 2008b, Mazzorana et al., 2014, Wierman et al., 2013, Cipriani et al., 2012). However, with the advent of hybrid pixel array detectors (Broennimann et al., 2006a) it is now possible to collect large enough numbers of datasets in a single synchrotron visit to allow robust statistical analysis techniques to be applied. Using this technology in the present study, formal statistical analysis was performed on both the control and HC1-transferred crystal populations, as described below.

The Mann-Whitney U-test was used to compare the distributions of the data quality indicators R_{merge}, $R_{p.i.m.}$, mean $I/\sigma(I)$, maximum high resolution limit, mosaicity and unit cell parameters between control and HC-transferred crystals. Data were analysed by testing the null hypothesis that data quality is the same between the two populations (i.e. remains unchanged by treatment in the humid air stream). The Mann-Whitney U-test is a nonparametric test so makes no assumptions about the defining properties of the data, such as being normally distributed (Mann and Whitney, 1947). However, nonparametric tests are less likely to reject a false null hypothesis and most of our data are normally, or at least weakly normally, distributed; so an independent sample two tailed Student's t-test assuming equal variance between populations was also conducted (Student, 1908). Both statistical tests were applied independently to all protein crystal types (further details are provided in Appendix:

Statistical Tests). Lysozyme and ferritin populations collected on beamlines I02 and I03 were analysed separately as crystals originated from different crystallisation batches.

2.3. Results

Owing to the fragile nature of protein crystals, excessive manipulation can have a negative effect on their physical integrity and the quality of diffraction data. To minimise the effects of manual handling, micromanipulators were used to carefully position the donor and acceptor mounts and to wick excess solvent without touching the protein crystal. This ensured the process was as gentle as possible on the crystals, while mechanical and osmotic stresses were kept to a minimum. During the transfer, none of the crystals showed signs of physical damage (breaking, cracking etc), dissolving or drying out while being transferred in the relative humidity stream. The measured relative humidities, using the loop method described in the earlier section "Defining Relative Humidity", are listed for each crystal type in tables 2.3-5. The averaged data collection parameters and their corresponding standard deviations, for both the control and HC-transferred samples, are presented for lysozyme (Table 2.3), insulin and thaumatin (Table 2.4) and ferritin (Table 2.4).

2.3.1. Crystal Integrity

Exposure to a relative humidity stream can have significant effects on the crystal lattice, with naked crystals being particularly susceptible to the effects of the humid air stream (Sanchez-Weatherby et al., 2009a). These effects have been demonstrated for various protein crystals where controlled dehydration (Chotiyarnwong et al., 2007, Bowler et al., 2006c, Takayama and Nakasako, 2011, Awad et al., 2013, Adachi et al., 2009, Bowler et al., 2007, Bowler et al., 2006a, Engel et al., 2003, Henrich et al., 2003, Kyrieleis et al.,

2005, Hu et al., 2011) and other more serendipitous methods (Abergel, 2004, Cramer et al., 2000, Esnouf et al., 1998, Gupta et al., 2010, Heras et al., 2003, Kuo et al., 2003, Nakamura et al., 2007, Vijayalakshmi et al., 2008, Yap et al., 2007) have been used to alter crystal packing and improve the diffracting power. Having determined the relative humidity with the loop method described above, the crystals should remain unchanged while in the humid air stream. To test the effects of the humid stream and wicking on the crystal lattice, average mosaicity (as defined by XDS (Kabsch, 2010b)) and unit cell parameters were analysed between treatments using the Mann-Whitney U-test and Student's t-test.

2.3.1.1. Unit Cell Parameters

Lysozyme 1, lysozyme 2, insulin and thaumatin all showed no significant difference in average unit cell parameters between control and HC-transferred crystals by the Mann-Whitney U-test (Table 2.6) or Student's t-test (Table 2.7). However, ferritin 1 and ferritin 2 showed a statistically significant increase (Table 2.6 and 2.7) in unit cell parameters of 0.34 Å and 0.61 Å (Table 2.5), respectively, following transfer in the humid air stream.

2.3.1.2. Mosaicity

The average mosaicity of lysozyme 1, lysozyme 2, insulin and ferritin 2 also show no significant difference between control and HC-transferred crystals. Although lysozyme 2 shows a statistically significant improvement in HC-transferred sample mosaicity, closer inspection of individual lysozyme 2 datasets reveals two control data sets with greatly increased mosaicity (Fig. 2.14). Analysis with the Grubb's test for outliers (Grubbs, 1950) determined the two control datasets to be significant outliers for mosaicity within the control population (circled in red in Figure 2.14). However, as their increased mosaicity has

no effect on the other data quality indicators under investigation they were kept in the control population. Both the Mann-Whitney U-test and Student's t-test reported a statistically significant difference in mosaicity for control and HC-transferred ferritin 1 data (Table 2.6 and 2.7). HC-transferred crystals showed an average decrease in mosaicity of 0.018° compared to the control samples.

For the thaumatin data, contradictory results were obtained with the Mann-Whitney U-test suggesting no significant difference and Student's t-test suggesting a significant improvement in mosaicity from the HC-transferred population. This discrepancy between the statistical tests may be the result of the small control population (7 crystals) and the control populations having almost double the variance of HC-transferred samples (Table 2.4), preventing accurate determination of the Z-value (Equation #3: Appendix). The spread of individual mosaicity between control and HC-transferred samples suggests HC-transferred samples have a lower mosaicity, with the control population spanning 0.041-0.097° while the HC-transferred population cover 0.029-0.071° (Fig. 2.16). Of the 14 HC-transferred samples, 7 are lower than or approximately equal to the lowest mosaicity value for the control population (Fig. 2.16). A plot of the average mosaicity for both thaumatin populations also demonstrates a clear difference between populations with HC-transferred samples having a lower mosaicity by well over one standard deviation (Fig. S.#). Based on this it stands to reason that Student's t-test is correct to identify these populations as significantly different.

2.3.2. Data Quality Indicators

The effects on data quality of both exposure to the humid air stream and the wicking to facilitate crystal transfer, were assessed by comparing the average data quality

indicators (R_{merge}, $R_{p.i.m.}$, mean I/σ(I), and maximum high resolution limit) between control and HC-transferred crystals. To quantify the differences/similarities between control and HC-transferred samples these individual data quality indicators have been plotted against mosaicity and correlation coefficients have been calculated. Average data quality indicators were also plotted against the average maximum high resolution limit, with standard deviations in horizontal and vertical directions.

2.3.2.1. Lysozyme

Individual lysozyme 1 data quality indicators are plotted against mosaicity in figure 2.13. All data quality indicators show a similar spread between treatments and the correlation coefficients (Table 2.A1) show a similar positive linear relationship between treatments, with increasing R_{merge}, $R_{p.i.m.}$ and high resolution limit with increasing mosaicity; and a negative linear relationship between decreasing I/σ(I) and increasing mosaicity. Average Lysozyme 1 data quality indicators calculated from 8 control samples and 12 HC-transferred samples (Table 2.3) showed no significant difference in overall or high resolution R_{merge}, $R_{p.i.m.}$, I/ σ(I) or maximum high resolution limit based on the Mann-Whitney U test or Student's t-test (Table 2.6 and 2.7). This is consistent with the data presented in Figure 2.19, where average values of all data quality indicators for both treatments are within one standard deviation in horizontal and vertical directions.

Lysozyme 2 data mirrors the patterns observed in lysozyme 1 populations. Similar distributions of control and HC-transferred data are observed (Fig 2.14) and correlation coefficients (Table 2.A1) agree with the positive and negative trends described above. Average lysozyme 2 data calculated from 18 control samples and 16 HC-transferred samples (Table 2.3) also show no statistically significant differences by either statistical test (Table

2.6 and 2.7), with both data points within on standard deviation of each other in vertical and horizontal directions (Fig. 2.19).

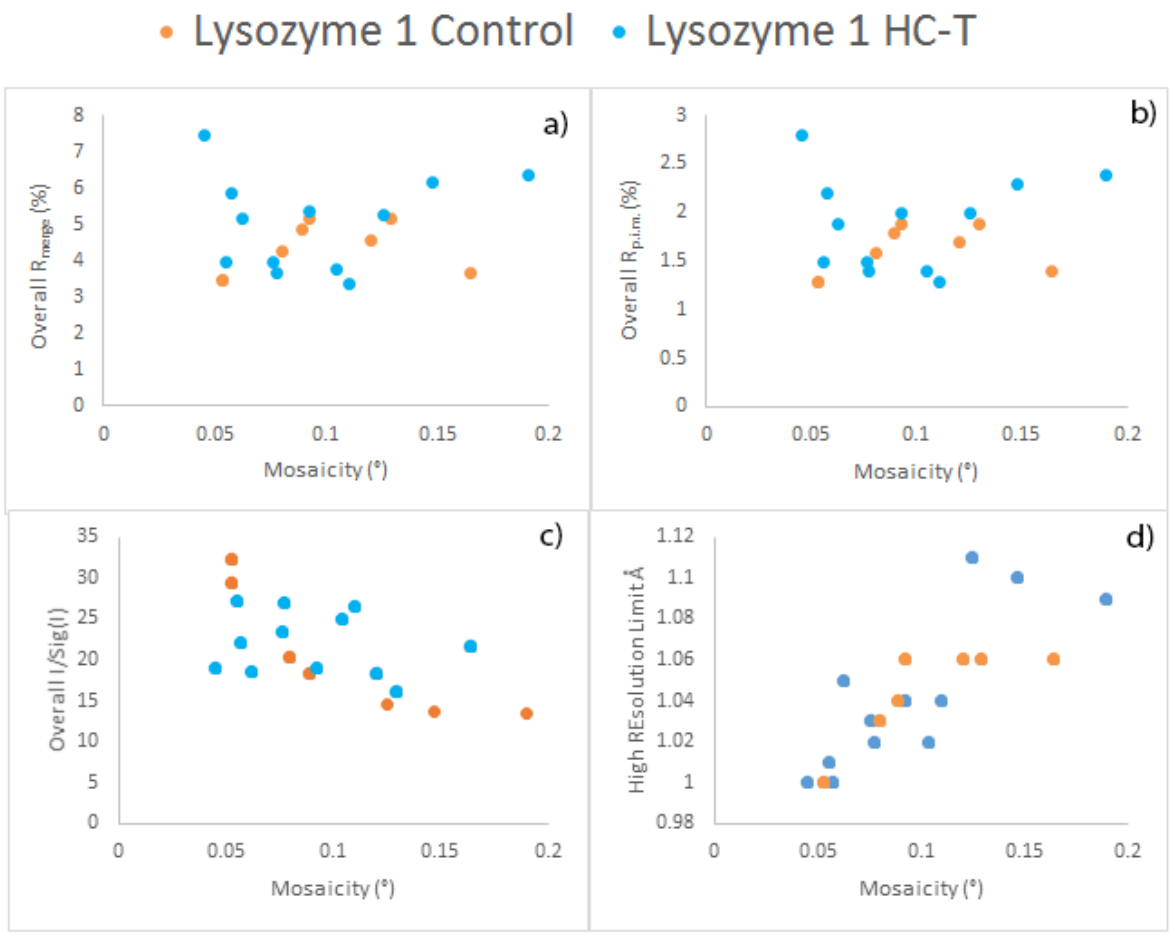


Fig. 2.13: Individual dataset overall data quality indicators plotted against mosaicity for lysozyme one control and HC-transferred samples at λ = 1.0 Å. (a) Overall R_{merge} (%). (b) Overall $R_{p.i.m.}$ (%). (c) Overall mean I/σ(I). (d) Maximum high resolution limit (Å).

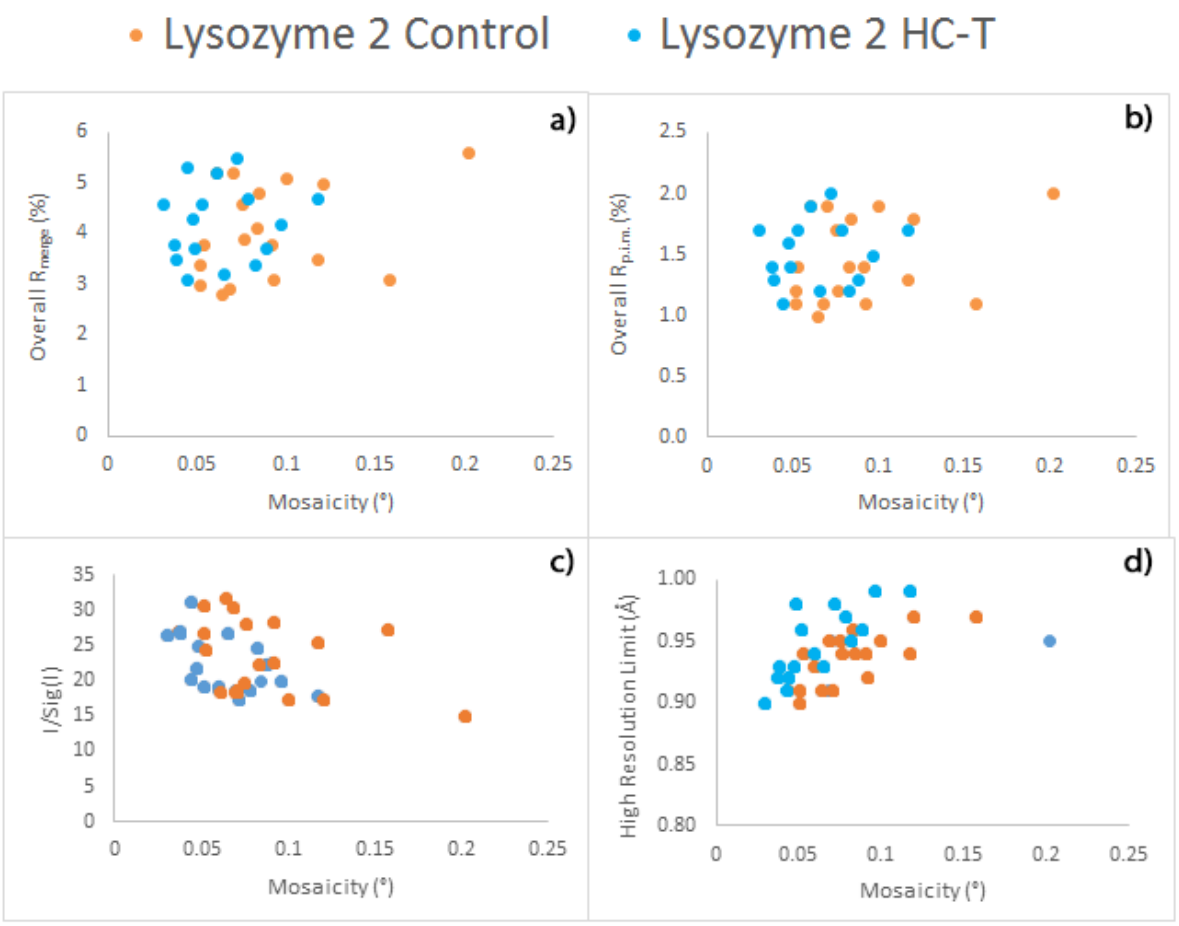


Fig. 2.14: Individual dataset overall data quality indicators plotted against mosaicity for lysozyme two control and HC-transferred samples at λ = 1.0 Å. (a) Overall Rmerge (%). (b) Overall Rpim (%). (c) Overall mean I/σ(I). (d) Maximum high resolution limit (Å).

Table 2.3: Average data-collection statistics for tetragonal lysozyme. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC* are shown as overall statistics followed by the highest resolution bin below.

Test Crystal	**Lysozyme 1**		**Lysozyme 2**	
Method	Control	HC-Transferred	Control	HC-Transferred
No. Crystals	8	12	18	16
Crystal Size (mm)	0.26 (0.02) x 0.24 (0.04) x 0.13(0.02)	0.28 (0.07) x 0.23 (0.04) x 0.16 (0.03)	0.23 (0.06) x 0.19 (0.03) x 0.16 (0.03)	0.23 (0.04) x 0.20 (0.03) x 0.14 (0.03)
Relative Humidity (%)	N/A	89	N/A	89
Space Group	$P4_12_12_1$	$P4_12_12_1$	$P4_12_12_1$	$P4_12_12_1$
Unit-cell Parameters (Å)	a = b = 78.95 (0.10), c = 36.94 (0.04)	a = b =78.99 (0.06), c = 36.96 (0.07)	a = b = 78.96 (0.08), c = 36.89 (0.04)	a = b = 78.97 (0.04), c = 36.91 (0.03)
Mosaicity (°)	0.098 (0.036)	0.095 (0.041)	0.09 (0.038)	0.06 (0.02)
Maximum High Resolution Limit (Å)‡	1.04 (0.03)	1.04 (0.04)	0.94 (0.02)	0.95 (0.03)
High Resolution Bin Range (Å) ▼	1.04-1.06	1.04-1.06	0.95-0.97	0.95-0.97
Multiplicity	7.3 (0.1)	7.3 (0.3)	8.4 (0.9)	8.3 (0.1)
	3.0 (0.2)	2.9 (0.3)	7.6 (0.6)	7.6 (0.5)
Completeness (%)	98.2 (0.8)	98.1 (1.0)	99.6 (1.0)	99.3 (1.5)
	81.8 (6.3)	81.4 (6.9)	99.3 (2.5)	98.1 (3.8)
R_{merge}[+] (%)	4.4 (0.7)	5.1 (1.2)	4.1 (0.9)	4.2 (0.7)
	43.3 (14.9)	48.4 (27.9)	81.2 (29.8)	91.2 (33.6)
$R_{p.i.m.}$* (%)	1.6 (0.2)	1.9 (0.5)	1.5 (0.4)	1.5 (0.3)
	27.4 (9.4)	31.8 (19.6)	31.0 (10.9)	34.8 (12.0)
$\langle \mathbf{I}/\boldsymbol{\sigma}(\mathbf{I}) \rangle$	22.0 (5.4)	20.8 (4.9)	23.5 (5.1)	22.7 (4.0)
	2.5 (1.2)	2.5 (1.2)	2.3 (0.7)	2.3 (0.9)
CC*[¥]	0.9991 (0.0003)	0.9986 (0.0010)	0.9995 (0. 0005)	0.9992 (0.0006)
	0.7821 (0.1082)	0.7418 (0.2049)	0.7680 (0.1187)	0.7377 (0.1346)

‡Maximum high resolution limit defined by AIMLESS using an I/σ(I) > 2 in the highest resolution shell. ▼Highest resolution bin based on I/σ(I) > 2. [+]$R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i(hkl)$ and *$R_{p.i.m.} = \sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i(hkl)$, where $I_i(hkl)$ represents observed intensity of the ith measurement of the reflection hkl and $\langle I(hkl)\rangle$ represent the mean intensity of reflection hkl after scaling. [¥]$CC^* = \sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}}$

2.3.2.2. Insulin

Individual insulin data quality indicators show comparable distributions when plotted against mosaicity (Fig 2.15), with correlation coefficients indicating similar trends with increasing mosaicity as observed in lysozyme 1 and 2 presented in the appendix (Table 2.A1). Average data quality indicators for insulin were calculated from 16 control and 13 HC-transferred crystals (Table 2.4). Overall R_{merge}, $R_{p.i.m.}$ and mean I/σ(I) all show significant differences between populations, with HC-transferred populations having significantly lower R_{merge} and $R_{p.i.m.}$ and higher mean I/σ(I) by both statistical tests (Table 2.6 and 2.7). This is clearly seen in figure 2.19a and 2.19c where overall insulin data points (royal blue) are over one standard deviation apart in the vertical direction. Despite this discrepancy in the overall

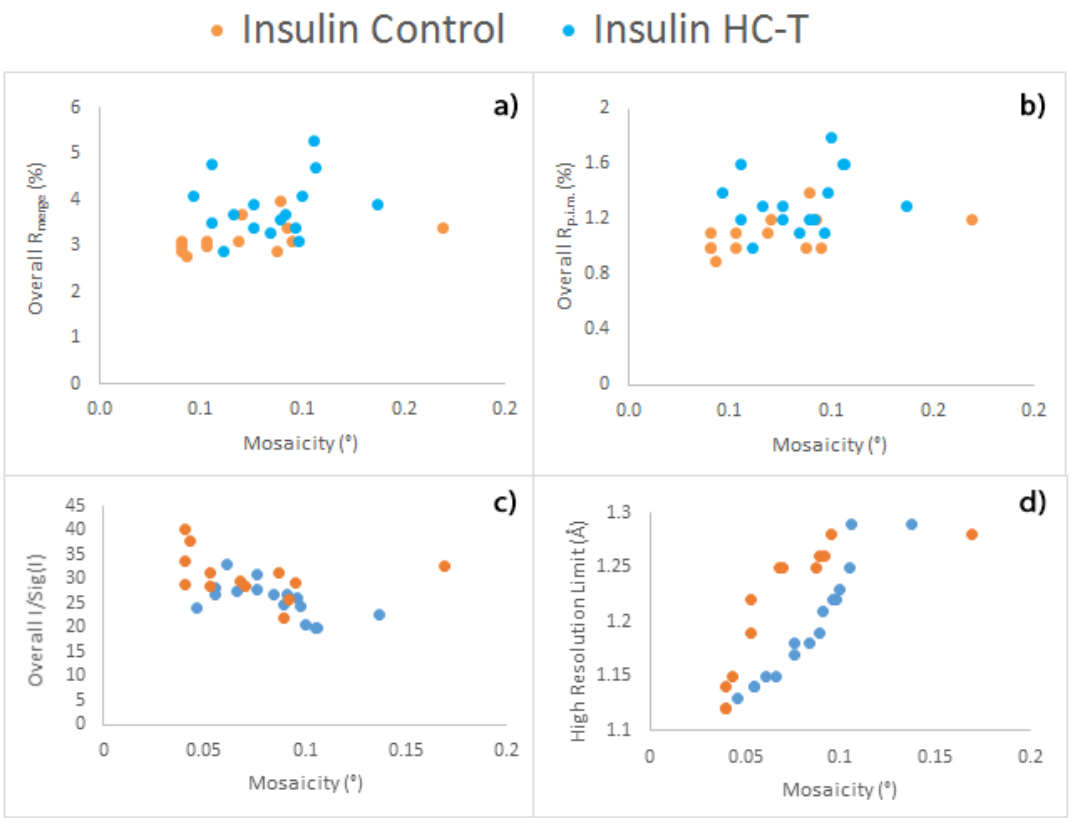


Fig 2.15: Individual dataset overall data quality indicators plotted against mosaicity for Insulin control and HC-transferred samples at λ = 1.0 Å. (a) Overall R_{merge} (%). (b) Overall $R_{p.i.m.}$ (%). (c) Overall mean I/σ(I). (d) Maximum high resolution limit (Å).

data, high resolution shell data quality indicators are statistically the same (Table 2.6 and 2.7). Analysis of low resolution shell data (~7-55 Å) show a statistically significant improvement in I/σ(I) in HC-transferred samples. This may explain the improvement in overall data while high resolution shell data remain unaffected. Despite this difference in overall data quality indicators the maximum high resolution limit between treatments show no significant differences (Table 2.6 and 2.7).

2.3.2.3. Thaumatin

Individual thaumatin data show distinct distribution of data quality indicators relative to mosaicity. A small group in the HC-transferred population presents with increased R_{merge} and $R_{p.i.m.}$ and decreased I/σ(I) towards the lower end of the mosaicity axis (Fig. 2.16a-c). This group skews the calculation of correlation coefficients for HC-transferred populations giving merging R factors a negative correlation and I/σ(I) a positive correlation with increasing mosaicity; contrary to all other data (Table 2.A1). Average data quality indicators for thaumatin were calculated from a control population of 7 crystals and a HC-transferred population of 14 crystals. No statistically significant differences in overall or high resolution data quality indicators were found by either statistical tests (Table 2.6 and 2.7).This correlates well with thaumatin data in figure 2.19, with all averages within one standard deviation in vertical and horizontal directions.

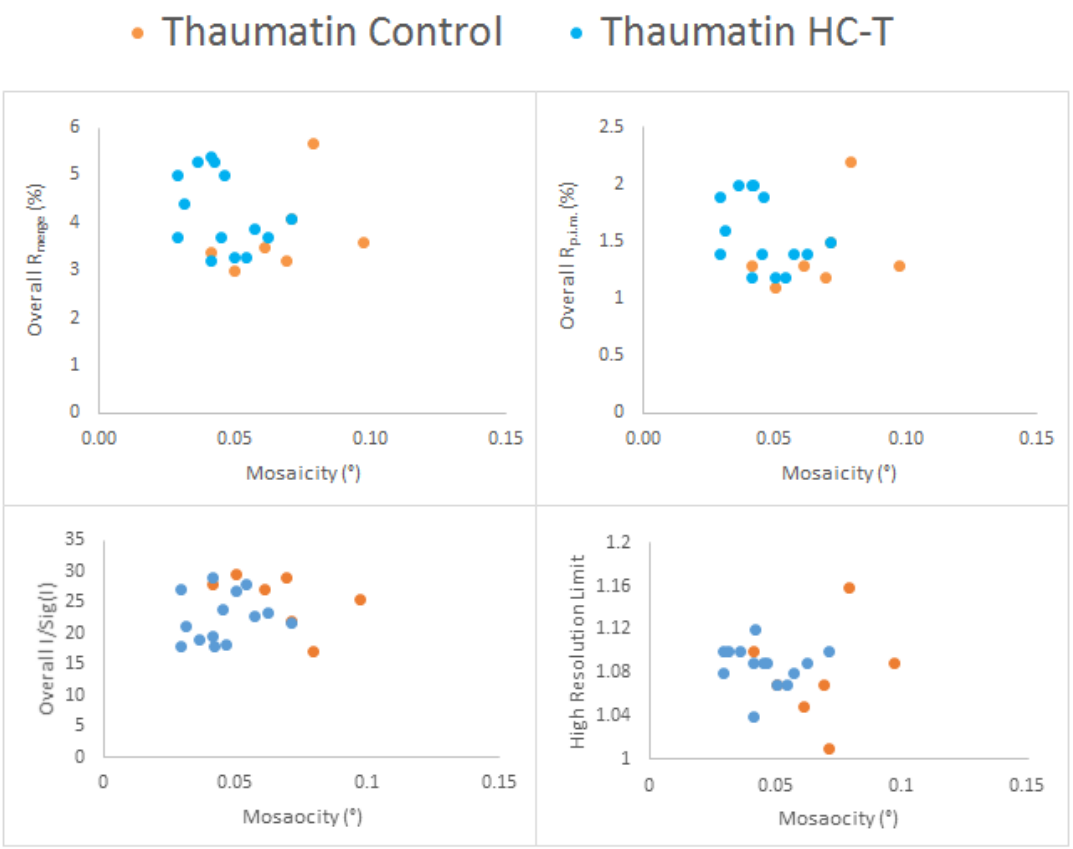


Fig 2.16. Individual dataset overall data quality indicators plotted against mosaicity for thaumatin control and HC-transferred samples at λ = 1.0 Å. (a) Overall R_{merge} (%). (b) Overall $R_{p.i.m.}$ (%). (c) Overall mean I/σ(I). (d) Maximum high resolution limit (Å).

Table 2.4: Average values for data collection and processing statistics of insulin and thaumatin crystals. Values in parentheses show standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below.

Test Crystal	**Insulin**		**Thaumatin**	
Method	Control	HC-Transferred	Control	HC-Transferred
No. Crystals	16	13	7	14
Crystal Size (mm)	0.17 (0.03) x 0.15 (0.03) x 0.12 (0.02)	0.17 (0.02) x 0.15 (0.02) x 0.13 (0.02)	0.20 (0.01) x 0.12 (0.02) x 0.11 (0.01)	0.20 (0.04) x 0.12 (0.02) x 0.10 (0.02)
Relative Humidity (%)	N/A	91	N/A	86
Space Group	$I2_13$	$I2_13$	$P4_12_12_1$	$P4_12_12_1$
Unit-cell Parameters (Å)	a = b = c 78.24 (0.13)	a = b = c = 78.29 (0.08)	a = b = 57.85 (0.03) c = 150.23 (0.06)	a = b = 57.84 (0.05) c = 150.27 (0.13)
Mosaicity (°)	0.084 (0.023)	0.072 (0.035)	0.067 (0.017)	0.045 (0.012)
Maximum High Resolution Limit (Å)‡	1.20 (0.05)	1.21 (0.06)	1.08 (0.04)	1.09 (0.02)
Highest Resolution Bin Range (Å) ▼	1.21-1.23	1.21-1.23	1.08 -1.10	1.08 -1.10
Multiplicity	9.6 (0.1)	9.6 (0.1)	7.5 (0.2)	7.4 (0.1)
	8.9 (0.1)	8.9 (0.1)	3.2 (0.5)	2.7 (0.5)
Completeness (%)	100.0 (0.0)	100.0 (0.0)	96.8 (1.9)	97.0 (4.0)
	100.0 (0.1)	100.0 (0.1)	74.0 (10.5)	77.6 (17.8)
R_{merge}⁺ (%)	3.8 (0.6)	3.2 (0.3)	3.8 (0.8)	4.2 (0.8)
	103.3 (51.4)	123.3 (57.2)	56.0 (33.8)	50.2 (16.9)
$R_{p.i.m.}$* (%)	1.3 (0.2)	1.1 (0.13)	1.4 (0.3)	1.6 (0.3)
	36.4 (18.1)	43.6 (20.3)	34.5 (20.4)	33.6 (9.8)
Mean I/σ(I)	26.0 (3.3)	30.7 (4.6)	25.5 (4.2)	22.7 (3.8)
	2.3 (1.1)	2.2 (1.4)	2.3 (0.9)	2.0 (0.6)
CC*¥	0.9994 (0.0005)	0.9998 (0.0004)	0.9997 (0.0005)	0.9995 (0.0005)
	0.7341 (0.1545)	0.6653 (0.1843)	0.7510 (0.2150)	0.7506 (0.1119)

•Refer to Table 2.3 for definitions

2.3.2.4. Ferritin

Individual ferritin 1 R_{merge}, $R_{p.i.m.}$ and high resolution limit show similar positive linear relationships with increasing mosaicity between treatments based on correlation coefficients (Table 2.A1), however; there is a clear skew towards lower mosaicity, R_{merge}, $R_{p.i.m.}$, high resolution limit and higher high resolution limit in HC-transferred samples (Fig. 2.17). Average data quality indicators for ferritin 1 were calculated from 16 control crystals and 19 HC-transferred crystals. Average overall and high resolution R_{merge}, $R_{p.i.m.}$ and high resolution limit show no significant difference between treatments, along with overall mean I/σ(I) (Table 2.6 and 2.7). However, high resolution mean I/σ(I) does show a significant improvement in HC-transferred samples. These differences and similarities can be seen in Figure 2.19.

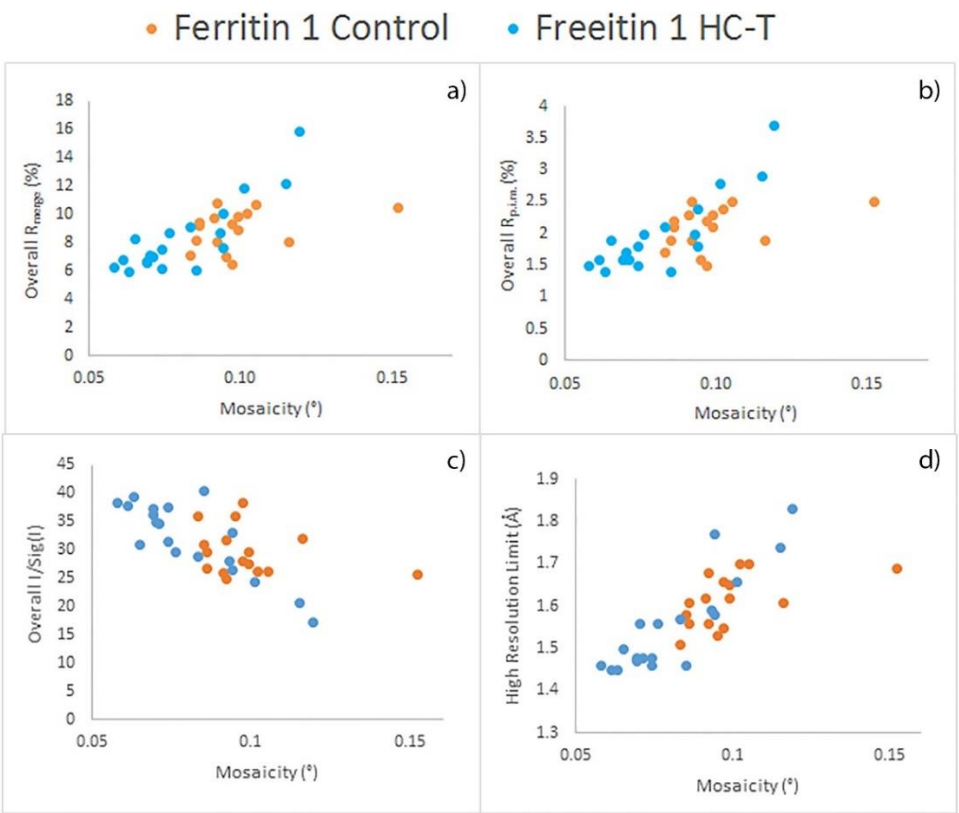


Fig. 2.17: Individual dataset overall data quality indicators plotted against mosaicity for ferritin one control and HC-transferred samples at λ = 1.0 Å. (a) Overall R_{merge} (%). (b) Overall $R_{p.i.m.}$ (%). (c) Overall mean I/σ(I). (d) Maximum high resolution limit (Å).

Individual ferritin 2 dataset data quality indicators present with a similar spread to ferritin 1 based on correlation coefficients (Table 2.A1). However, the skew towards lower mosaicity seen in ferritin 1 is not observed in ferritin 2; with both control and HC-transferred samples clustered close together (Fig 2.18). Average overall and high resolution data quality indicators for ferritin two show no significant differences by both statistical tests and all data points in figure 2.19 (light blue) are within one standard deviation in horizontal and vertical directions.

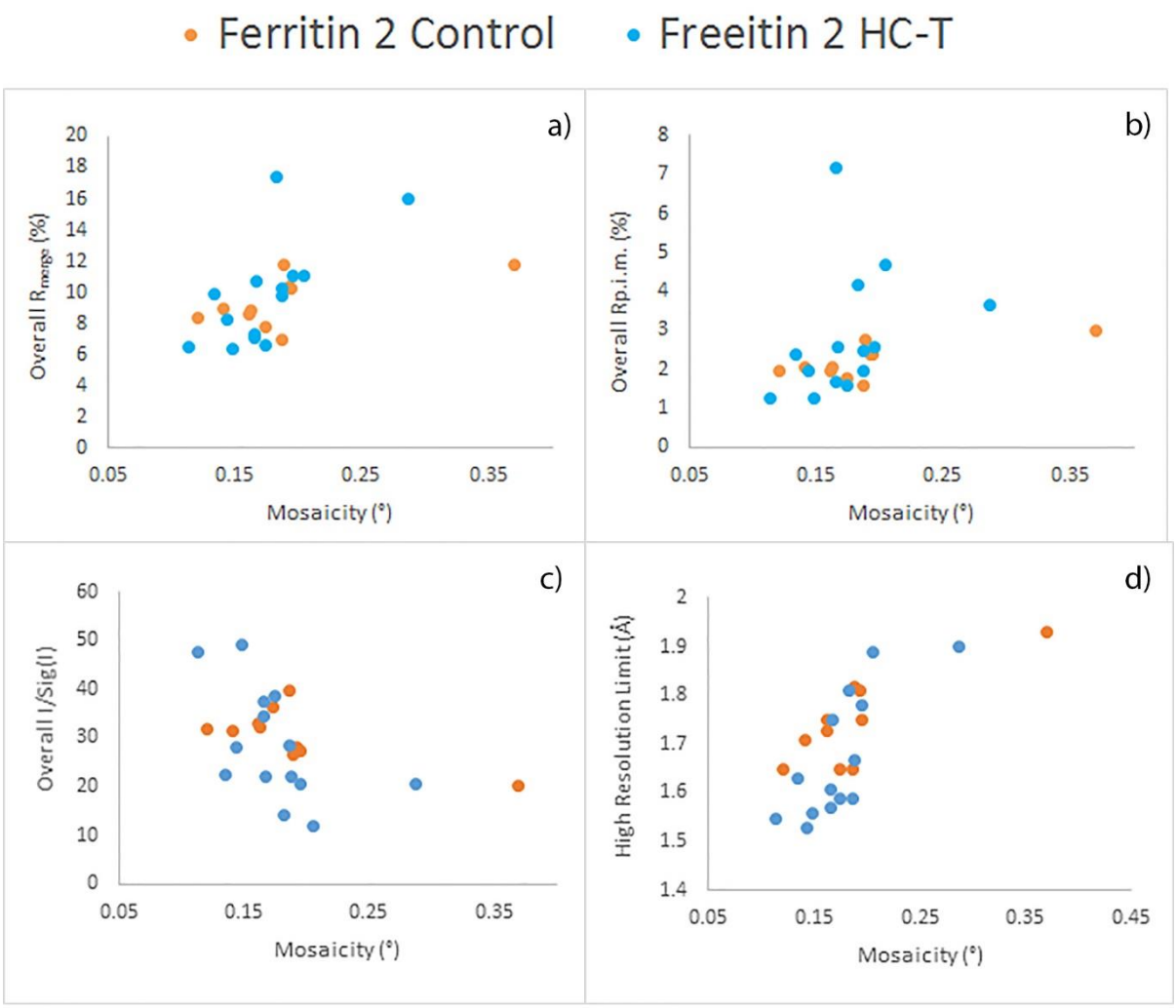


Fig. 2.18: Individual dataset overall data quality indicators plotted against mosaicity for ferritin two control and HC-transferred samples at λ = 1.0 Å. (a) Overall R_{merge} (%). (b) Overall $R_{p.i.m.}$ (%). (c) Overall mean I/σ(I). (d) Maximum high resolution limit (Å).

Table 2.5: Average data-collection statistics for Cubic ferritin. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin.

Test Crystal	**Ferritin 1**		**Ferritin 2**	
Method	Control	HC-Transferred	Control	HC-Transferred
No. Crystals	16	19	10	14
Crystal Size (mm)	0.09 (0.01) x 0.08 (0.01) x 0.08 (0.01)	0.084 (0.006) x 0.076 (0.006) x 0.078 (0.012)	0.15 (0.02) x 0.14 (0.01) x 0.12 (0.01)	0.15 (0.02) x 0.14 (0.01) x 0.12 (0.02)
Relative Humidity (%)	N/A	89.5	N/A	90.0
Space Group	*F*432	*F*432	*F*432	*F*432
Unit-cell Parameters (Å)	a = b = c = 181.37 (0.17)	a = b = c = 181.71 (0.14)	a = b = c = 181.13 (0.17)	a = b = c = 181.74 (0.32)
Mosaicity (°)	0.099 (0.006)	0.081 (0.017)	0.189 (0.064)	0.175 (0.023)
Maximum High Resolution Limit (Å)‡	1.61 (0.06)	1.56 (0.11)	1.75 (0.09)	1.67 (0.09)
Highest Resolution Bin Range (Å) ▼	1.75-1.78	1.75-1.78	1.80 -1.84	1.80 -1.84
Multiplicity	19.0 (0.1)	19.0 (0.1)	18.9 (0.6)	19.4 (2.9)
	19.3 (0.0)	19.3 (0.1)	19.3 (1.7)	20.2 (2.7)
Completeness (%)	100.0 (0.0)	99.9 (0.3)	100.0 (0.1)	99.4 (0.7)
	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)	100.0 (0.1)
R_{merge}⁺ (%)	9.0 (1.3)	8.4 (2.5)	9.4 (1.6)	9.9 (2.9)
	98.2 (24.3)	75.4 (41.4)	131.9 (61.1)	108.3 (34.7)
$R_{p.i.m.}$* (%)	2.1 (0.3)	2.0 (0.6)	2.2 (0.4)	2.5 (0.8)
	22.8 (5.7)	17.9 (9.9)	31.5 (17.4)	26.1 (8.2)
Mean I/σ(I)	29.8 (4.0)	32.1 (6.4)	30.9 (5.2)	28.6 (10.6)
	3.9 (1.1)	5.6 (2.0)	3.0 (1.1)	4.3 (1.7)
CC*¥	1.00 (0.00)	1.00 (0.00)	0.9999 (0.0003)	0.9992 (0.0006)
	0.88 (0.05)	0.92 (0.07)	0.8103 (0.1292)	0.8686 (0.0532)

•Refer to Table 2.3 for definitions

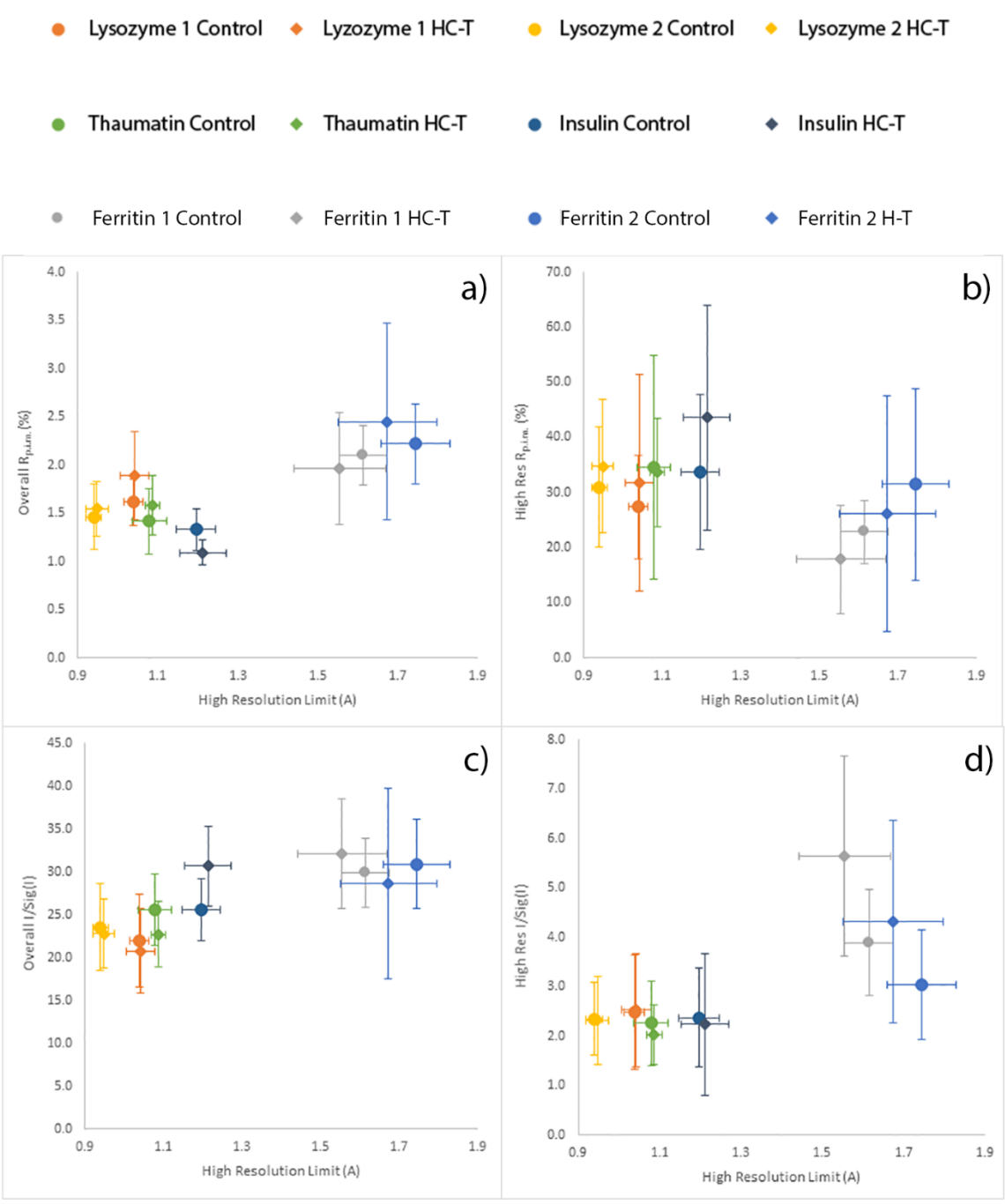


Fig. 2.19: Average lysozyme 1, lysozyme 2, thaumatin, insulin, ferritin 1 and ferritin 2 data quality indicators for control (circles) and HC-transferred (diamonds) populations collected at λ = 1 Å plotted against average maximum high resolution limit. Error bars represent one standard deviation. (a) Overall $R_{p.i.m.}$. (b) High resolution $R_{p.i.m.}$. (c) Overall mean I/σ(I). (c) High resolution mean I/σ(I).

Table 2.6: Mann-Whitney U-test results for all data quality indicators. + represents populations that are statistically the same and - represents populations that are statistically different. Symbols in parentheses are for data in the highest resolution bin ranges defined in Table 2.3-5.

Test Crystal	Lysozyme 1	Lysozyme 2	Insulin	Thaumatin	Ferritin 1	Ferritin 2
Unit Cell Parameter (a=b=) (Å)	+	+	+	+	-	-
Unit Cell Parameter (c=) (Å)	+	+	N/A	+	N/A	N/A
Mosaicity (°)	+	-	+	+	-	+
R_{merge}⁺ (%)	+ (+)	+ (+)	- (+)	+ (+)	+ (+)	+ (+)
$R_{p.i.m.}$* (%)	+ (+)	+ (+)	- (+)	+ (+)	+ (-)	+ (+)
I/σ(I)	+ (+)	+ (+)	- (+)	+ (+)	+ (-)	+ (+)
Maximum High Resolution Limit (Å) ‡	+	+	+	+	-	+

‡Maximum high resolution limit defined by AIMLESS using an I/σ(I) > 2 in the highest resolution shell. ▼Highest resolution bin based on I/σ(I) > 2. ⁺R_{merge} = $\sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i(hkl)$ and *$R_{p.i.m.}$ = $\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i(hkl)$, where $I_i(hkl)$ represents observed intensity of the ith measurement of the reflection hkl and $\langle I(hkl)\rangle$ represent the mean intensity of reflection hkl after scaling.

Table 2.7: Student's t-test results for all data quality indicators from an independent two-tailed t-test assuming equal variance between populations. + represents populations that are statistically the same and - represents populations that are statistically different. Symbols in parentheses are for data in the highest resolution bin ranges defined in Table 2.3-5.

Test Crystal	Lysozyme 1	Lysozyme 2	Insulin	Thaumatin	Ferritin 1	Ferritin 2
Unit Cell Parameter (a = b =) (Å)	+	+	+	+	-	-
Unit Cell Parameter (c =) (Å)	+	+	N/A	+	N/A	N/A
Mosaicity (°)	+	-	+	-	-	+
R_{merge}⁺ (%)	+ (+)	+ (+)	- (+)	+ (+)	+ (+)	+ (+)
$R_{p.i.m}$*	+ (+)	+ (+)	- (+)	+ (+)	+ (+)	+ (+)
I/σ(I)	+ (+)	+ (+)	- (+)	+ (+)	+ (-)	+ (+)
Maximum High Resolution Limit (Å)	+	+	+	+	+	+

‡Maximum high resolution limit defined by AIMLESS using an I/σ(I) > 2 in the highest resolution shell. ▼Highest resolution bin based on I/σ(I) > 2. ⁺R_{merge} = $\sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i(hkl)$ and *$R_{p.i.m.}$ =

$\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_i|I_i(hkl)-\langle I(hkl)\rangle|/\sum_{hkl}\sum_i(hkl)$, where $I_i(hkl)$ represents observed intensity of the ith measurement of the reflection hkl and $\langle I(hkl)\rangle$ represent the mean intensity of reflection hkl after scaling.

2.4. Discussion

This work aimed to assess the use of a tuneable humid air stream to preserve macromolecular crystals in preparation for long wavelength X-ray crystallography. Datasets from large populations of control and HC-transferred crystals were collected and subjected to robust statistical tests, with a view to assess any detrimental effects caused by exposure to the humid air stream and/or extra manual handling to remove excess solvent. Macromolecular crystals, being typically ~50% solvent (Matthews, 1968, Matthews, 1976), are sensitive to changes in their relative humidity and can result in changes to crystal packing (Russi et al., 2011). Lysozyme, insulin, thaumatin and ferritin have solvent contents of 30%, 26%, 57% and 60%, respectively. To monitor crystal packing between control and HC-transferred samples, changes in unit cell parameters and mosaicity were assessed, while assuming that these parameters and crystal packing would remain constant at the correct relative humidity (Sjogren et al., 2002, Sjogren and Hajdu, 2001b, Sjogren and Hajdu, 2001a). In these circumstances we would expect diffraction data quality to be unaffected by HC-transfer at the 'standard' X-ray crystallography data collection wavelengths (~1 Å), where X-ray absorption is negligible (Arndt, 1984, Polikarpov et al., 1997).

2.4.1. Crystal Integrity

The average unit cell parameters of lysozyme 1, lysozyme 2, thaumatin and insulin remained constant between treatments, and are comparable with published data (Murray and Garman, 2002, Lopez-Jaramillo et al., 2001, Meents et al., 2007); suggesting the correct relative humidity was measured by the loop method, and mechanical/osmotic stresses were

kept to a minimum. The same can be said for the mosaicity of lysozyme 1, lysozyme 2 (after taking the outliers into account) and insulin. However, thaumatin crystals showed a statistically significant decrease in mosaicity following HC-transfer. The exact reasons behind this apparent increase in order remains unknown, but may stem from more efficient flash cooling by removing surrounding solvent, or exposure to the humidity controlled stream altering the solvent content of crystals. Ferritin 1 and 2 both showed a statistically significant increase in unit cell parameters following HC-transfer. Relative humidity tests by the loop method gave a relative humidity of 89.5% and 90% for ferritin 1 and 2, respectively. An increase in unit cell parameters may result from exposure to too high a relative humidity, leading to incorporation of water, which increased in volume upon plunge cooling in liquid nitrogen. Although the effects of this increase in unit cell parameters on data quality indicators was negligible, discussed in the following section, it suggests the correct relative humidity for ferritin was not determined. For future experiments a higher resolution microscope may be necessary to observe the small changes in drop size for more accurate relative humidities. Ferritin crystals may also be sensitive to the oscillating nature of the humid air stream which moves approximately ± 0.5% around the set relative humidity; as the packing essentially forms a hollow sphere which may contract or expand with rising and falling humidity.

2.4.2. Data Quality indicators

Overall and high resolution shell data quality indicators show no significant differences between treatments for lysozyme 1, lysozyme 2, thaumatin, ferritin 1 and ferritin 2, with the exception of high resolution mean I/σ(I) in ferritin 1, which shows a significant improvement in HC-transferred samples. High resolution I/σ(I) is highly

dependent on the resolution limit. Although HC-transferred samples were found to have a higher resolution limit than control samples (1.56 Å and 1.61 Å, respectively) this difference was not found to be statistically significant. Reduced background scatter as a result of removing excess solvent may have improved the signal-to-noise ratio leading to a stronger high resolution I/σ(I); however, if this were the case it would likely be seen in other test crystals.

Although there was no statistically significant effect on the overall averages between HC-transferred and control populations, a small population of datasets with lower mosaicity, increased R_{merge} and $R_{p.i.m.}$ and decreased mean I/σ(I) was observed in HC-transferred thaumatin crystals (Fig 2.16). It has been reported that lower mosaicity crystals are more susceptible to gas-induced loop motion caused by movement of flexible sample mounts in the cryo stream, which can have detrimental effects on data quality (Alkire et al., 2008, Thorne et al., 2003, Alkire et al., 2013). This effect is enhanced by removing excess solvent, which reduces the stiffness of a sample mount (Alkire et al., 2013). This would explain the subset of poor quality thaumatin crystals following HC-transfer. Despite this poor quality subset as a result of low mosaicity there were no statistically significant differences between treatments as the large population studied accounted for these outliers.

Overall data quality indicators in HC-transferred insulin crystals also showed a significant improvement over control samples. Low resolution I/σ(I) significantly increases in HC-transferred samples but a similar decrease in low resolution R_{merge} and $R_{p.i.m.}$ is not observed despite being related to the intensities of a reflection (Karplus and Diederichs, 2012, Weiss, 2001, Arndt et al., 1968). The exact nature of this improvement in low

resolution and overall data quality following HC-transfer is unknown. Improvements could stem from a decrease in background scatter and absorption by reducing surrounding solvent, as discussed previously (Kiefersauer et al., 2000, Kitago et al., 2005a, Kitago et al., 2010a). Improvements in data quality could also be the result of more efficient flash cooling after removing surrounding solvent, or by exposure to the humidity controlled stream altering the solvent content of crystals.

2.5. Conclusions

As modern MX is addressing increasingly challenging projects, there is a greater need to improve the quality of diffraction data. This is of particular importance with weak signals such as native anomalous diffraction using long wavelength MX or weak diffraction from microcrystals. A key aspect of improving these weak signals is to reduce the background scatter and absorbance during data collection. Here a method to remove the solvent surrounding protein crystals with a view to aid S-SAD or P-SAD phasing using long wavelength X-rays, where achieving a low background absorption is critical has been outlined, tested and validated. Although designed for future long wavelength MX experiments (for example, using I23 at Diamond), the Crystal transfer method described here may also find applications in the rapidly developing field of X-ray free electron laser crystallography (Boutet et al., 2012). Techniques such as serial crystallography are gaining momentum (Gati et al., 2014), and could greatly benefit from the tailored mounting of macromolecular crystals onto specially designed supports. Here I have demonstrated that diffraction quality from macromolecular crystals can be preserved, and in some cases enhanced, at short (‘standard’) wavelengths by manipulation in a humid air stream and removal of excess solvent by wicking. Slight changes in mosaicity and unit cell parameters in

some protein crystals suggests the technique may benefit from a more accurate means of determining relative humidity, but overall protein crystals appear unaffected by exposure to the humid air stream. Comparison of statistics from refined models was considered, however, it was not performed due to time constraints and the lack of a systematic method for model building to minimise random errors. The next step of this process, described in the following chapters, is to apply this method and the robust statistical analysis to more challenging macromolecular targets and to long wavelength data collection.

2.6. Appendix

2.6.1. Example Scripts

2.6.1.1. XSCALE.INP

```
!MAXIMUM_NUMBER_OF_PROCESSORS=16
!RESOLUTION_SHELLS= 10 6 4 3 2.5 2.0 1.8 1.7 1.6
!SPACE_GROUP_NUMBER=19
!UNIT_CELL_CONSTANTS=65.46 108.41 113.15   90.000  90.000  90.000
!REIDX=-1 0 0 0    0 -1 0 0    0 0 -1 0
!REFERENCE_DATA_SET= fae-rm.ahkl

!MINIMUM_I/SIGMA=3.0
!REFLECTIONS/CORRECTION_FACTOR=50   !minimum #reflections/correction_factor
!0-DOSE_SIGNIFICANCE_LEVEL=0.10
!WFAC1=1.5 ! factor applied to e.s.d.'s before testing equivalent reflections
!SAVE_CORRECTION_IMAGES= FALSE ! TRUE is default

 OUTPUT_FILE=datasetX_Y.HKL                  !at minimum of f'
 FRIEDEL'S_LAW=FALSE !TRUE
 MERGE=FALSE !TRUE
 STRICT_ABSORPTION_CORRECTION=TRUE  !FALSE is default
 INPUT_FILE= .../datasetX_Y/XDS_ASCII.HKL
!    INCLUDE_RESOLUTION_RANGE= 20 1.6
!    CORRECTIONS= DECAY MODULATION ABSORPTION
!    CRYSTAL_NAME=Seleno1 !Remove first "!" to switch on 0-dose extrapolation
!    STARTING_DOSE=0.0  DOSE_RATE=1.0  !Use defaults for 0-dose extrapolation
```

2.6.1.2. XSCALE and Pointless

```
module load ccp4
module load XDS
module load global/cluster

cd Dataset1_1
xscale
pointless -copy xdsin dataset1_1.HKL hklout pointless.mtz
echo "done"
cd ..
cd dataset1_2
xscale
pointless -copy xdsin dataset1_2.HKL hklout pointless.mtz
echo "done"
cd ..
```

2.6.1.3. Aimless

```
module load ccp4
module load XDS
module load global/cluster

cd dataset1_1
aimless hklin pointless.mtz hklout aimless.mtz << eof > aimless.log
run 1 all
scales constant
ANALYSIS isigminimum 2
resolution 60.0 0.00
anomalous on
eof
cd ..
cd dataset1_2
aimless hklin pointless.mtz hklout aimless.mtz << eof > aimless.log
run 1 all
scales constant
ANALYSIS isigminimum 2
resolution 60.0 0.00
anomalous on
eof
cd ..
```

2.6.1.4. Data Extraction

```
echo dataset1_1 hc1 > Protein_Wavelength_Results
grep "                                                    Overall    InnerShell    OuterShell" dataset1_1/ aimless.log
>>Protein_Wavelength_Results
grep "Low resolution limit" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "High resolution limit" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Rmerge  (within I+/I-)" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Rmerge  (all I+ and I-)" dataset1_1/ aimless.log >> 1A_EAL_results
grep "Rmeas (within I+/I-)" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Rmeas (all I+ & I-)" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Rpim (within I+/I-)" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Rpim (all I+ & I-)" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Rmerge in top intensity bin" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Total number of observations" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Total number unique" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Mean((I)/sd(I))" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Mn(I) half-set correlation CC(1/2)" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Completeness                    " dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Multiplicity                   " dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Anomalous completeness              " dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Anomalous multiplicity             " dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "DelAnom correlation between half-sets" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "Mid-Slope of Anom Normal Probability" dataset1_1/ aimless.log >> Protein_Wavelength_Results
grep "REFLECTING_RANGE_E.S.D.=" dataset1_1/XDS_ASCII.HKL >> Protein_Wavelength_Results
grep "UNIT_CELL_CONSTANTS=" dataset1_1/XSCALE.INP >> Protein_Wavelength_Results
echo >> Protein_Wavelength_Results
echo >> Protein_Wavelength_Results
```

2.6.2. Statistical Tests

2.6.2.1. Mann-Whitney U-test

All data from both populations were ranked in ascending order and assigned a rank of one to n, where n represents the total number of crystals from combined populations. Adding up the ranks for a population gives the ranked sum R_1, which was used to calculate U_1 using equation (1),

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad (2.1)$$

where n_1 and n_2 represent the number of samples in population one and two respectively. U_1 represents the number of times a data point from population one is smaller than a data point from population two, when comparing all possible pairs of samples, where one sample is taken from each population. U_1 was used to calculate U_2 using equation (2),

$$U_2 = n_1 n_2 - U_1$$

(2.2)

where U_2 represents the equivalent of U_1 for population two. In larger sample sizes (n_1 and n_2 > 10) ranked data from the Mann-Whitney U-test become near normally distributed. Using equation (3),

$$Z = \frac{2U - n_1 n_2}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{3}}}$$

(2.3)

we can transform the U statistic (the higher value of U_1 and U_2) to give a sampling distribution comparable to the standard normal (Z) distribution. Observed Z-values were compared to the critical Z-value, representing data ±1.96 standard deviations from the calculated mean (95% of the sample population). The null hypothesis was rejected if calculated Z-values exceeded the critical Z-value in either direction.

2.6.2.2. Student's t-test

The t-values were calculated using equation (1.4),

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(2.4)

where $\overline{X_1}$ is the mean of population one, $\overline{X_2}$ is the mean of population two and n_1 and n_2 are the number of samples in population one and two, respectively. S_p^2 represents the pooled sample variance, as defined by equation (5),

$$S_p^2 = \frac{df_1 S_1^2 + df_2 S_2^2}{df_1 + df_2}$$

(2.5)

where df_1 and df_2 represent the degrees of freedom of populations one and two (equations (6) and (7), respectively) and $S_1^2$and S_2^2 represent the sample variances of populations one and two, respectively. Calculated t-values were analysed against critical t-values from Student's t-distribution tables for α(2) = 0.05 (Student, 1908) with degrees of freedom calculated using equation (8),

$$df_1 = n_1 - 1$$

(2.6)

$$df_2 = n_2 - 1$$

(2.7)

$$DF = df_1 + df_2$$

(2.8)

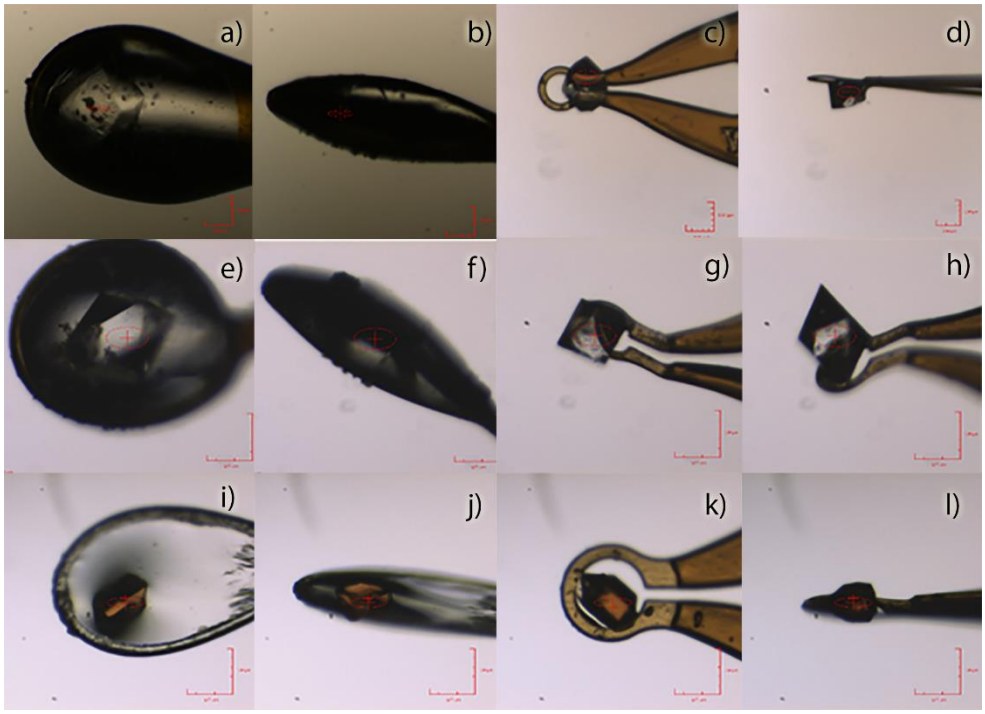
If observed t-values were found to be greater than the critical t-value, the null hypothesis is rejected. In addition to *t*-values, *P*-values were calculated using standard statistical analysis software. The *P*-value quantifies the uncertainty of obtaining the observed data if the null hypothesis were true. A critical *P*-value of 0.05 was used to judge significance, with the null hypothesis rejected if the observed *p*-values were less than or equal to the critical *P*-value.

2.6.3. Correlation Coefficients

Table 2.A1: Correlation coefficients of overall data quality indicators in relation to mosaicity.

Protein	**Lysozyme 1**		**Lysozyme 2**		**Insulin**		**Thaumatin**		**Ferritin 1**		**Ferritin 2**	
Treatment	Ctrl	HC	Ctrl	HC	Ctrl	HC	Ctrl	HC	Ctrl	HC	Ctrl	HC
R_{merge} (%) Vs Mosaicity (°)	0.29	0.14	0.32	0.11	0.17	0.44	0.42	-0.36	0.33	0.86	0.60	0.67
$R_{p.i.m.}$ (%) Vs Mosaicity (°)	0.32	0.15	0.30	-0.16	0.21	0.47	0.37	-0.39	0.36	0.86	0.68	0.34
I/σ(I) Vs Mosaicity (°)	-0.65	-0.58	-0.41	-0.56	-0.60	-0.29	0.51	0.17	-0.27	-0.84	-0.67	0.54
High Resolution Limit (Å) Vs Mosaicity (°)	0.85	0.82	0.58	0.77	0.95	0.78	0.15	-0.08	0.51	0.88	0.80	0.77

Chapter 3 – Excess Solvent and Data quality at Long Wavelength



3.1. Introduction

Data collection for phasing using long wavelengths (>2Å), including S-SAD, requires very accurate measurement of weak signals that can easily be swamped by more prominent signals and background absorption/scattering (Chapter 1). The weak anomalous signals of native sulphur atoms can be greatly enhanced by using X-ray wavelengths approaching the absorption edge energy of sulphur, ~ 5 Å. However, at longer wavelengths absorption of X-rays by air, the crystal, surrounding solvent and the sample mount increase along with background scatter. This makes accurate measurement of anomalous signals difficult. In an effort to reduce absorption by air, helium paths have been used (Biou et al., 2005) and in-vacuum crystallography is currently under development at the Diamond Light Source for beamline I23 (Wagner et al., 2015). As discussed in Chapter 2, a wide variety of sample or crystal handling techniques may be exploited (e.g. to remove excess surrounding solvent) to minimise the effects of absorption and background scatter, and this special handling is particularly applicable for long wavelength or S-SAD phasing experiments.

The 'optimum' wavelength for data collection has been a lively topic of discussion within the X-ray crystallography community for several years. It was thought that use of short (λ = 0.5 Å) and ultra-short (λ = 0.3 Å) wavelength X-rays could be used to reduce background scatter and absorption errors (Arndt, 1984). This idea was abandoned after it became apparent that there was no significant improvement in the signal-to-noise ratio compared to synchrotron collection at λ = 0.9 Å (Polikarpov et al., 1997). Although collection at ultra-short wavelengths was not attempted here, the results presented in Chapter 2 suggest that absorption and background scatter are not a limiting factor in data quality at 1 Å, even in the presence of excess surrounding solvent. Another consideration in the 'optimum' wavelength for data collection is crystal size. Recent work suggests that smaller crystals (≤ 10 µm) collected with micro-focus beams at high energies (40 keV) allow photoelectric escape due to increased photoelectron path length, thereby reducing energy

deposition within the crystal and minimising radiation damage (Cowan and Nave, 2008, Nave and Hill, 2005, Schneider, 2008, Holton and Frankel, 2010).

The possible significance of surrounding solvent on data quality has been discussed in previous studies using one of two single crystals (Cipriani et al., 2012, Kiefersauer et al., 2000, Kitago et al., 2010b, Kitatani et al., 2009, Wierman et al., 2013) but the true effects on data quality have never been systematically tested or quantified. Following the successful use of the HC1 to preserve crystals for removal of mother liquor (Chapter 2), this study investigates the effects of excess solvent on data quality for long wavelength data collection. To achieve this, the same sample handling techniques and statistical analyses described in Chapter 2 were applied to three different types of soluble protein crystals: bacteriocin syringacin M (BSM), *PA*3825-EAL and bacteriocin pectocin M2 (BPM2).

BSM is a colicin M-like bacteriocin from *Pseudomonas syringae* pv. *Tomato* DC3000 which kills susceptible cells by targeting lipid II phosphoester bonds for cleavage to inhibit peptidoglycan synthesis (Grinter et al., 2012, El Ghachi et al., 2006, Harkness and Braun, 1989). These colicin-like bacteriocins have potential for use as protein based antibiotics that target close relatives to the bacteriocin producing strain (Riley, 1998). The specificity of receptor binding in colicins is regulated through the central domain, which is flanked by an unstructured translocation domain at the N-terminus (Fig. 3.1A, Orange) and a cytotoxic domain at the C-terminal end (Fig 3.1A, Red). BSM has a compact mixed β-sheet/helix structure with an unstructured translocation domain typical of colicin-like bacteriocins (Grinter et al., 2012). The cytotoxic domain is comprised of a half β-barrel fold (Fig.3.1B, Red) unique to colicin M type bacteriocins (Zeth et al., 2008) which coordinates the active site metals.

Much like BSM, BPM2 is a colicin M-like bacteriocin with the half β-barrel fold (Fig. 3.2, Blue). However, BPM2 lacks the unstructured translocation domain seen in BSM and has a ferrodoxin domain (Fig. 3.2, Red) connected to the C-terminal cytotoxic domain (Fig. 3.2, Blue) by a

helical linker (Fig. 3.2, Green) (Grinter et al., 2014). Pectobacterium releases pectinases and cellulases to breakdown plant cell tissues and cause cell lysis; releasing intracellular nutrients which are taken up by the bacteria, of which ferrodoxin is the major iron containing protein (Grinter et al., 2012). Bacteriocin pectocin M2 allows pecterobacterium to obtain iron directly from plant ferrodoxins through interactions with TonB-dependent receptors (Grinter et al., 2014). Both BSM and BPM2 represent good examples of a biologically interesting targets for protein crystallographers, as well as a challenge for sample handling due to the high solvent content of both protein crystals, ~ 80% and ~ 70% respectively for BSM and BPM2 (Fig. 3.3). BPM2 presents a particular challenge for long wavelength MX due to the presence of a highly absorbing 2Fe-2S cluster.

*PA*3825-EAL is an EAL domain containing protein involved in the cleavage of cyclic-di-GMP, a ubiquitous signalling molecule in bacteria involved in regulation of biofilm formation. PA3825-EAL will be discussed in greater depth in Chapter 6.

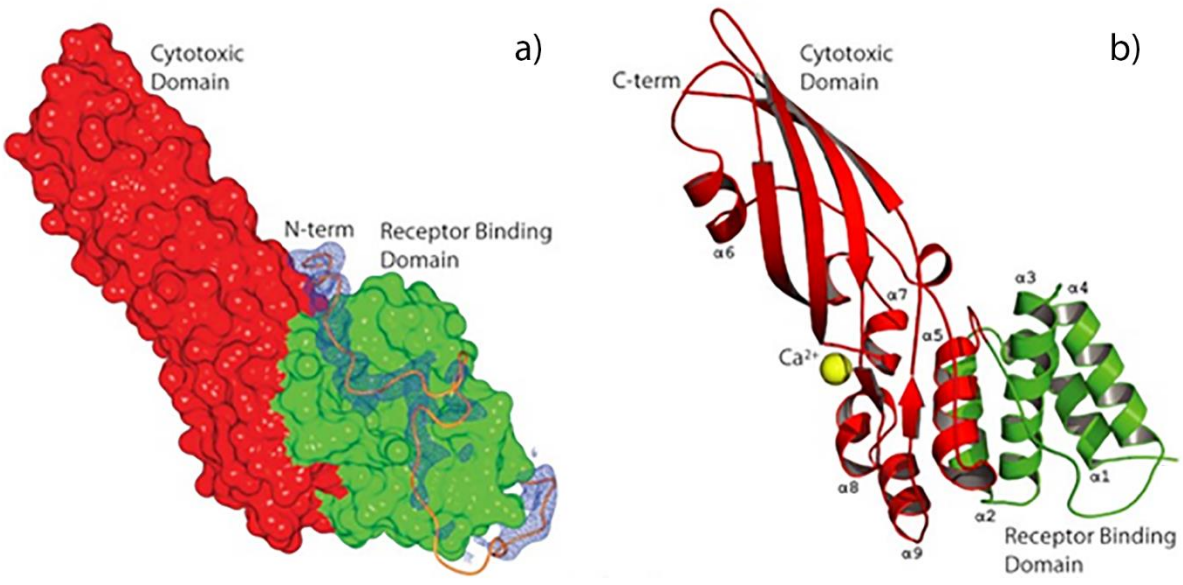


Fig. 3.1: The Structure of bacteriocin syringacin M. a) Surface model of the cytotoxic domain (red), receptor binding domain (green) and the unstructured N-terminus (residues 2-37) as a ribbon model with corresponding Fo-Fc omit map positive density contoured to 3 σ. b) Cartoon representation of bacteriocin syringacin M with domains coloured as in (a). The calcium ion is shown as a yellow sphere. Figure adapted from (Grinter et al., 2012).

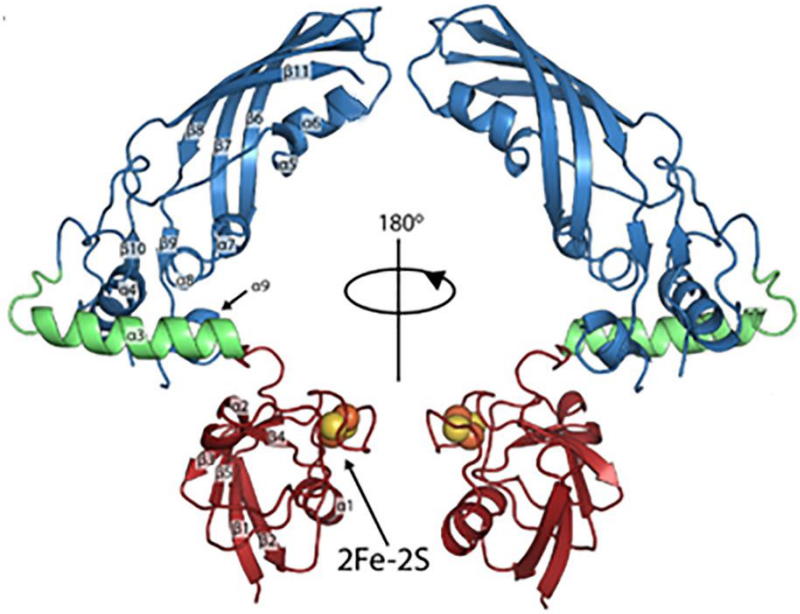


Fig. 3.2: The Structure of bacteriocin pectocin M. Cartoon representation with the cytotoxic domain in blue, plant like ferrodoxin domain in red and the linker helix in green. Iron and sulphur atoms from the 2Fe-2S cluster are represented by orange and yellow spheres, respectively. Figure adapted from (Grinter et al., 2014).

3.2. Methods

3.2.1. Protein Purification

Purified Bacteriocin Syringacin M (BMS) and Bacteriocin Pectocin M2 (BPM2) proteins were provided by Dr Rhys Grinter from the University of Glasgow and stored at -80°C (Grinter et al., 2014, Grinter et al., 2012). Details of PA3825-EAL expression and purification can be found in Chapter 6.

3.2.2. Crystallisation and Cryoprotection

All protein crystals were produced via the vapour diffusion method as either sitting or hanging drops. BSM protein was equilibrated against 8% w/v PEG 8000, 30% v/v ethylene glycol, 0.13 M $CaCl_2$, 0.03 M $MgCl_2$ and 0.1 M bicine/tris base pH 8.5 at a ratio of 2:3 in 5 µl hanging drops at 19°C. BSM crystals are grown in 30% ethylene glycol so required no additional cryo protection. BPM2 crystals were produced by equilibrating BPM2 proteins against 1.8 M ammonium sulphate, 3% MPD and 0.1 M Mes pH 6.5 at a 1:1 ratio in 5 µl sitting drops at 19°C. BPM2 crystals grew as large clusters

which were separated into single crystals with an acupuncture needle. Crystals were cryoprotected by transferring to a drop of reservoir solution with 25% (v/v) glycerol before flash cooling. *PA*3825-EAL crystal trays were set up using a Mosquito LCP liquid handling robot (TTP Labtech, Hertfordshire, England, UK). 15-20 mg/ml PA3825-EAL was equilibrated against 0.8 M sodium phosphate monobasic, 1.2 M potassium phosphate dibasic and 0.1 M sodium acetate pH 4.5 (final pH measured as pH 7.1) at a 1:1 ratio in 5 µl sitting drops at 4°C. *PA*3825-EAL crystals were cryoprotected by transferring crystals into 0.1 M sodium acetate pH 4.5 and 2.4 M zinc chloride before flash cooling in liquid nitrogen.

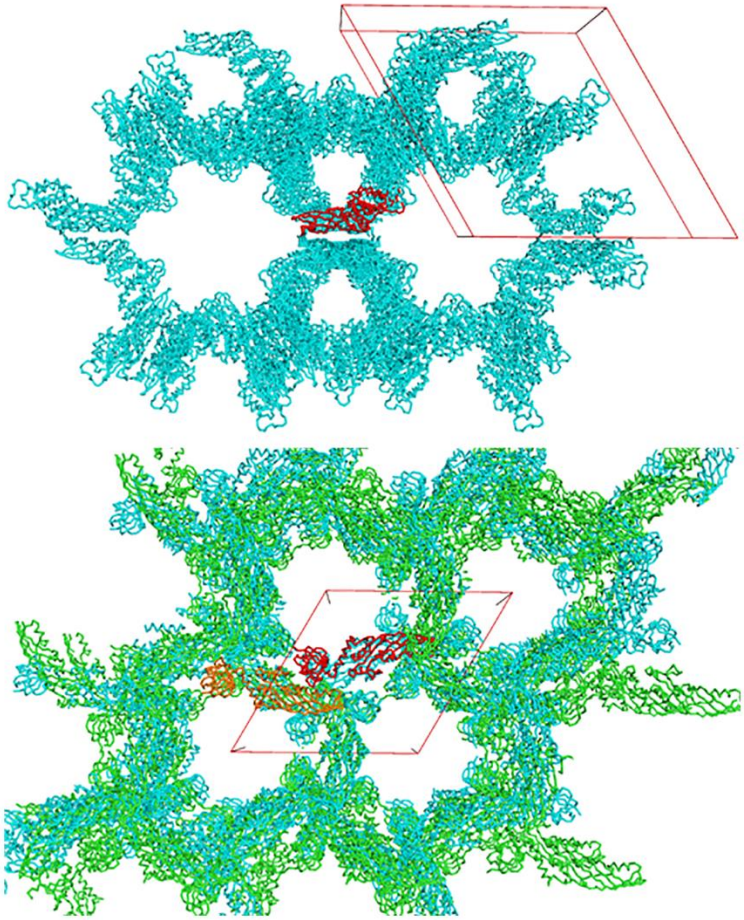


Fig. 3.3: Crystal packing of bacteriocin syringacin M (4FZM) in $P6_322$ and bacteriocin pectocin M2 (4N58) in $P3_121$. a) Crystal packing of bacteriocin syringacin M with symmetry mates (cyan) generated from the original structure (red) to 100 Å. The unit cell is shown in red . Solvent content 70%. b) Crystal backing of bacteriocin pectocin M2 with symmetry mates for chain A and B (green and cyan, respectively) generated from the original structure A and B chains (orange and red, respectively). The unit cell is shown in red. Solvent content 80%

3.2.3. Sample Preparation, Data Collection, Data Processing and Statistical Analysis

BSM, PA3825-EAL and BPM2 samples were prepared by the same protocol detailed in Chapter 2. Examples of control and HC-transferred samples for all proteins are presented in Figure 3.4. As PA3825-EAL and BMP2 crystals were not grown in their cryoprotectant, the relative humidity used for their transfer was determined from their cryoprotectant solutions not reservoir solutions. Crystals were thus transferred to their cryoprotectant solution before placing them in the humid air stream.

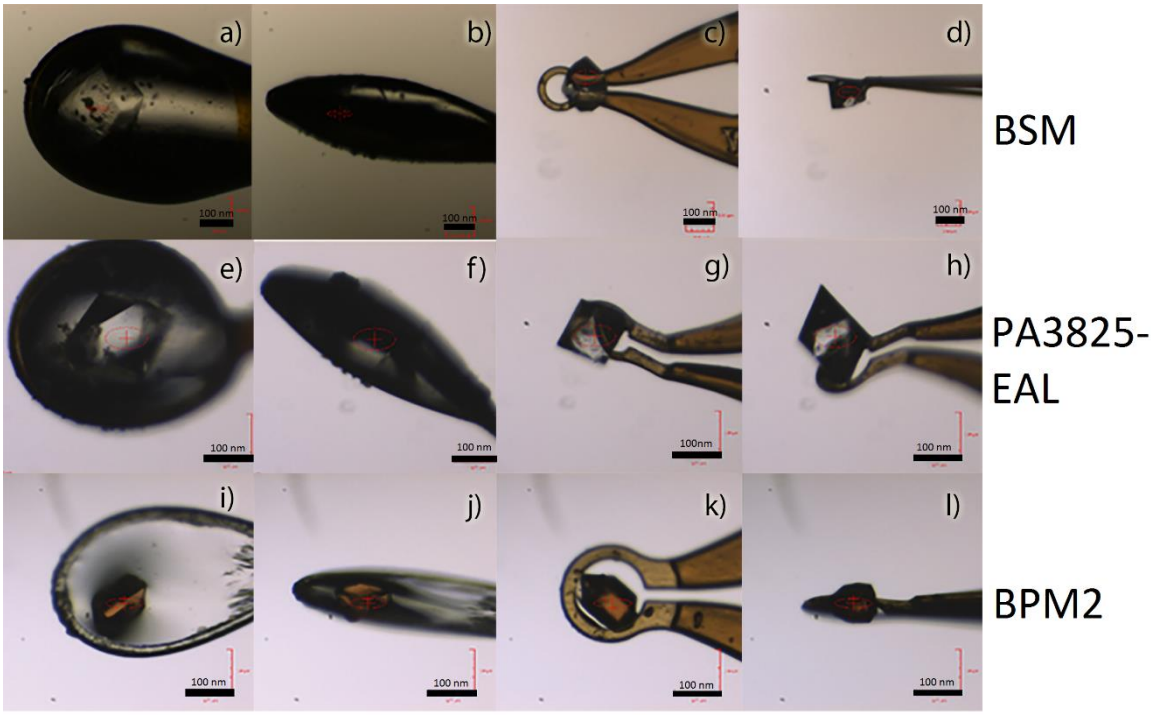


Fig. 3.4: Mounted control (left) and HC-transferred (right) protein crystals for bacteriocin syringacin M (a-d), PA3825-EAL (e-h) and bacteriocin pectocin M2 (i-l)

Data collection was performed as before, at 1 Å wavelength and with additional control and HC-transferred datasets collected at longer wavelengths (up to 2 Å). *PA*3825-EAL and BPM2 data were collected using the standard rotation/oscillation mode. BSM data were collected using the standard rotation data collection mode ('BSM-one') and also by the inverse beam method ('BSM-two'), to maximise the potential anomalous signal from the native sulphur atoms. The long

wavelength limit was determined from the maximum high resolution limit of the crystals and the shortest available distance between the crystal and the detector. To achieve this, three test images were taken at phi values 0°, 45° and 90° to give an indexing solution and a high resolution limit. Data collection parameters for all test crystals are summarised in table 3.1. All data were processed using the pipeline and bash scripts previously described, to determine the maximum high resolution limit and obtain data quality indicators (see Chapter 2). Control and HC-transferred populations at standard (λ = 1 Å) and longer wavelengths (λ ~ 2 Å) were compared to judge the effects of excess surrounding solvent on data quality at these wavelengths. In the case of BPM2, λ = 1 Å and λ = 2 Å controls and λ = 1 Å and λ = 2 Å HC-transferred populations were analysed for significant differences using a two tailed Student's t-test as described in Chapter 2.

Table 3.1: Data collection parameters for all crystals

Crystal	**BSM-1**		**BSM-2**		***PA*3825-EAL**		**BPM2**	
Wavelength (Å)	0.9795	2.0	1.0	1.8	0.9795	1.8	1.0	2.0
Beamline	I02		I04		I04		I03	
Beam Size (µm)	82 x 21		90 x 45		90 x 45		80 x 20	
Detector Distance (mm)	444.11	156.46	398.80	169.49	405.61	169.81	455.08	173.88
Maximum Resolution* (Å)	2.22		2.09		2.10		2.35	
Total Angular Range (°)	120		120		120		180	
Inverse Beam	No		Yes		No		No	
Space Group	$P6_322$		$P6_322$		$P4_32_12$		$P3_121$	
Transmission (%)	20		30		1		10	

All crystals were collected on a Pilatus 6M detector with a 0.1° step and 0.1 s exposure times. *Maximum resolution as defined by the detector edge at wavelengths and detector distances stated above.

3.2.4. RADDOSE-3D Analysis

All protein crystal types were analysed with RADDOSE-3D (Zeldin et al., 2013) between λ = 1.0 Å and λ = 2.4 Å to provide the elastic yield (photons), diffraction efficiency (photons/MGy), diffraction weighted dose (MGy) and the max dose (MGy). The flux for each wavelength at the appropriate transmission was calculated from the measured flux at Diamond Light Source I04 with the 100 µm aperture, 300 mA ring current and 100% transmission (Fig. 3.5) (www.diamond.ac.uk/Beamlines/Mx/i04). Theoretical values were calculated for crystals and buffer using the input parameters presented in Table 3.A2. Values for the buffer alone were calculated by changing the number of molecules in the asymmetric unit to 0 and increasing the solvent content to 100%. BPM2 RADDOSE-3D calculations were performed on two crystal shapes, with a long x-axis in crystal 1 (Fig. 3.6a) and a long z-axis in crystal 2 (Fig. 3.6b) as BPM2 crystals were rectangular in shape as opposed to more cubic for BSM and PA3825-EAL.

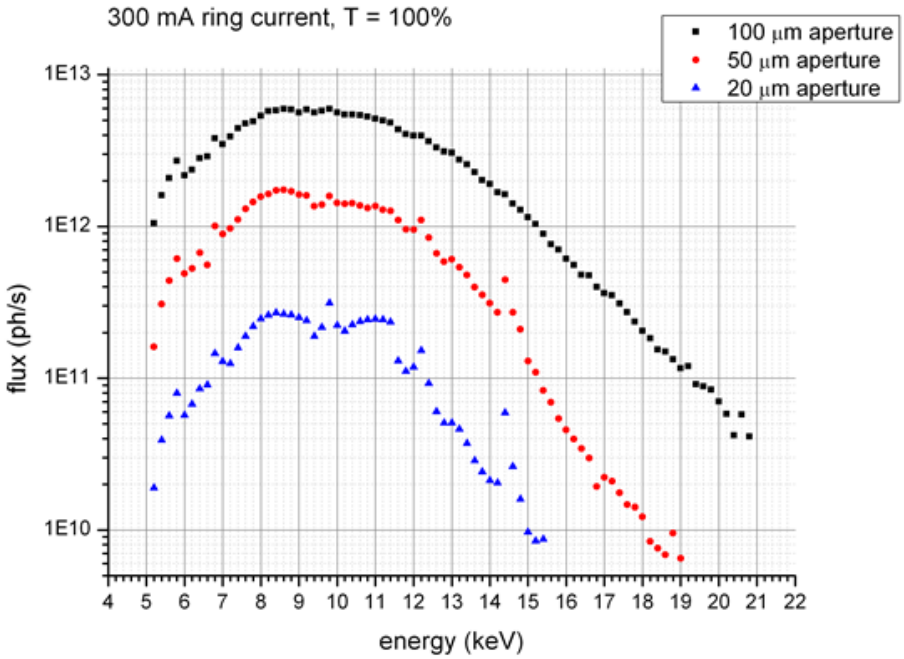


Fig. 3.5: Measured photon flux at different energies on Diamond light Source beamline I04 with a 300 mA ring current and 100% transmission (T) from January 2013 to December 2014 with different beam shaping apertures, 100 µm (black squares), 50 µm (red circles) and 20 µm as (blue triangles). Figure taken from www.diamond.ac.uk/Beamlines/Mx/i04 (14/05/2015).

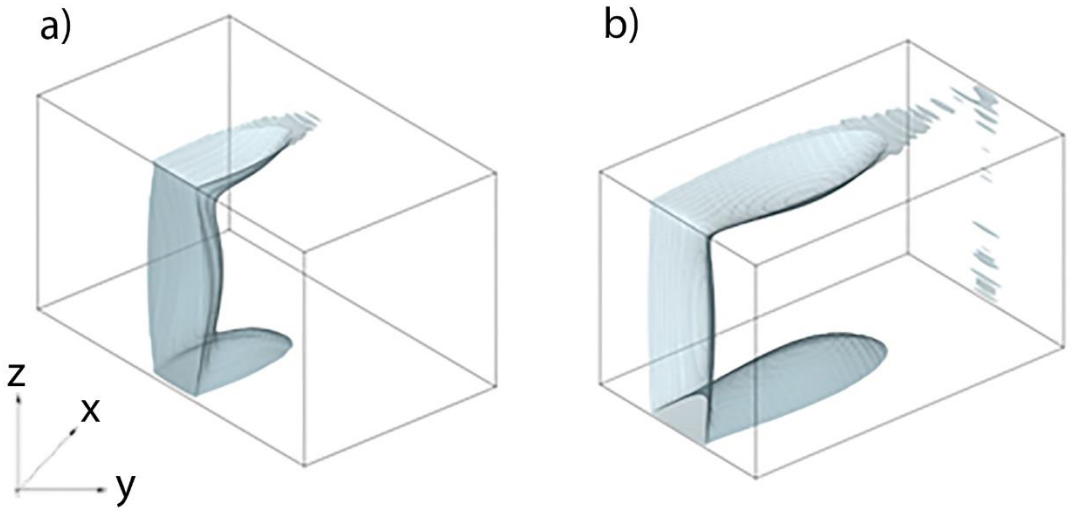


Fig. 3.6: A representation of Bacteriocin pectocin M2 crystals as they may be orientated in the X-ray beam, with orientation one (a) and two (b) as used in RADDOSE-3D simulations (Zeldin et al., 2013). Shaded areas represent the dose isosurface, a visual representation of deposited dose over the collected wedge (180°).

3.3. Results

As described in Chapter 2, diffraction data obtained from the crystals were assessed for crystal integrity based on the unit cell parameters and mosaicity; and estimates of diffraction data quality were based on R_{merge}, $R_{p.i.m.}$, $I/\sigma(I)$ and maximum high resolution limit, with statistical treatment using Student's t-test. All data were found to be normally or weakly normally distributed (Table 3.A3) so the Mann-Whitney U-test was not performed.

3.3.1. Crystal Integrity

BSM-one HC-transferred populations showed a consistent improvement in mosaicity compared to control populations (Table 3.6) at both wavelengths used. This difference is evident in the increased spread in mosaicity of control samples in figure 3.7. Unit cell parameters are also variable, with the c-axis significantly increasing in HC-transferred samples compared to control (Table 3.6). However, the a and b unit cell parameters remain constant between treatments (Table 3.6).

BSM-two control and HC-transferred populations at λ = 1 Å and λ = 1.8 Å both showed consistent a and b unit cell parameters. However, a statistically significant shrinkage in the c-axis was observed in the λ = 1 Å control population that was not observed in the λ = 1.8 Å control or HC-transferred populations (Table 3.6). The same pattern was observed for mosaicity, with the control population collected at λ = 1 Å exhibiting significantly higher mosaicity than all other populations. This can be seen in figure 3.8 where individual λ = 1 Å control and HC-transferred populations showed a very different distribution from each other, with the HC-transferred population having much lower mosaicity (0.04-0.15°) and shorter spread. On the other hand both treatments collected at λ = 1.8 Å showed a very similar distribution and spread of values for mosaicity (Fig. 3.8).

*PA*3825-EAL showed a statistically significant increase in average unit cell parameters (a and b) in HC-transferred populations, with an average increase of 0.20 Å and 0.14 Å for λ = 1 Å and λ = 1.8 Å populations, respectively. A significant increase in the c axis was also observed in the control population at λ = 1 Å, while no significant differences were seen between λ = 1.8 Å populations (Table 3.6). Mosaicity remained constant between all treatments collected at λ = 1 Å and λ = 1.8 Å, as seen in figure 3.10. Both treatments at λ = 1 Å showed some significant outliers, as identified by the Grubb's test (Grubbs, 1950). These datasets have been circled in red in figure 3.10 and were excluded from calculations of averages, standard deviations and the statistical analysis.

BPM2 crystals also showed a statistically significant reduction in unit cell parameters (a and b) in the HC-transferred populations at λ = 1 Å and λ =2 Å (Table 3.6), while unit cell parameters in the c-axis remain constant, as judged by Student's t-test (Table 3.6). A statistically significant decrease in mosaicity was observed in control populations at both wavelengths (Table 3.6). This is evident in Figure 3.12, particularly in control datasets collected at λ = 1 Å where all data points are clustered around 0.07°. One outlier was identified in the λ = 1 Å control population by the Grubb's test and is circled in red in figure 3.12. Again, this data point was excluded from averages, standard deviations and statistical tests.

3.3.2. Data Quality Indicators

3.3.2.1. Bacteriocin syringacin M-one

Individual BSM-one data quality indicators for λ = 1 Å and λ = 2 Å are plotted against mosaicity in figure 3.7a-h. In data collections at λ = 1 Å and λ = 2 Å, HC-transferred samples show more consistent R_{merge}, $R_{p.i.m.}$ and I/σ(I) while control samples show a much greater spread; with higher R-factors (Fig. 3.7a-d), lower I/σ(I) (Fig. 3.7e-f) and lower high resolution limit (Fig. 3.7g-h).

Average data quality indicators were calculated for 7/7 and 7/8 control/HC-transferred crystals for λ = 1 Å and λ = 2 Å, respectively, and are presented in Table 3.2. The λ = 1 Å populations show no difference in overall R_{merge}, $R_{p.i.m.}$ or I/σ(I), however, high resolution statistics are significantly improved in HC-treated samples in all cases, including the high resolution limit (Table 3.6). This is evident in figure 3.14 where average control data (grey circles) and HC-transferred data (grey diamonds) are within one standard deviation in the y axis in overall plots (Fig. 3.14a and c), and well over one standard deviation apart in the y axis in high resolution plots (Fig. 3.14b and d). Data collected at λ = 2 Å show a significant improvement in all data quality indicators following HC-transfer with the exception of average high resolution I/σ(I) (Table 3.6). Again, this is evident in figure 3.14 where only average high resolution I/σ(I) treatment populations are within one standard deviation of the other.

3.3.2.2. Bacteriocin Syringacin M-two

Individual BSM-two data quality indicators for λ = 1 Å and λ = 1.8 Å are plotted against mosaicity in figure 3.8a-h. All λ = 1A data show a clear trend of lower overall R_{merge}, R_{pim} and mosaicity, higher I/σ(I) and high resolution limit following exposure to the humid air stream. However, both populations show a similar trend of increasing R_{merge} and R_{pim} and decreasing I/σ(I) and high resolution limit with increasing mosaicity (Fig. 3.8a, c, e and g). Datasets collected at λ = 1.8

Å have a much more consistent spread in mosaicity between control and HC-populations, but HC-transferred populations still show an improvement in all data quality indicators (Fig. 3.8b, d, f and h).

The differences mentioned above are reflected in the Student's t-test results (Table 3.6), where λ = 1 Å HC-transferred R_{merge}, $R_{p.i.m.}$, I/σ(I) and high resolution limit are significantly better than controls. λ = 1.8 Å HC-transferred populations show similar significant improvements with the exception of high resolution limit, which remains constant between treatments. Average data quality indicators were calculated from 8/10 and 10/10 control/HC-transferred crystals for λ = 1 Å and λ = 1.8 Å, respectively, and are presented in Table 3.3. Average data quality indicators plotted against average high resolution limit in Figure 3.14 clearly show the difference between treatments; with all data quality indicators (except λ = 1.8 Å high resolution limit) over one standard deviation apart.

Table 3.2: Average data-collection statistics for bacteriocin syringacin M-one crystals at λ = 1 Å and λ = 1.8 Å. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below.

Crystal	**Bacteriocin Syringacin M One λ = 1 Å**		**Bacteriocin Syringacin M One λ = 2.0 Å**	
Method	Control	HC-Transferred	Control	HC-Transferred
No. Crystals	7	7	7	8
Crystal Size (mm)				
Relative Humidity (%)	N/A	89.0	N/A	89.0
Space Group	*P*6$_3$22	*P*6$_3$22	*P*6$_3$22	*P*6$_3$22
Unit-cell Parameters (Å)	a = b = 161.09 (0.50) c = 99.26 (0.35)	a = b = 160.76 (0.47) c = 99.73 (0.01)	a = b = 161.1 (0.20) c = 99.37 (0.11)	a = b = 161.00 (0.31) c = 99.76 (0.26)
Mosaicity (°)	0.137 (0.065)	0.064 (0.017)	0.194 (0.039)	0.151 (0.033)
Maximum High Resolution Limit (Å)‡	3.02 (0.05)	2.88 (0.07)	3.05 (0.13)	2.94 (0.05)
High Resolution Bin Range (Å) ▼	3.26-3.05	3.26-3.05	3.26-3.05	3.26-3.05
Multiplicity	12.2 (1.8)	12.1 (2.6)	11.5 (1.8)	12.5 (0.2)
	12.3 (1.8)	12.2 (2.7)	11.6 (1.8)	12.3 (0.1)
Completeness (%)	99.7 (0.3)	98.6 (1.8)	99.5 (0.2)	98.9 (1.4)
	100.0 (0.0)	99.2 (1.4)	99.9 (0.2)	99.7 (0.6)
R_{merge}⁺ (%)	8.8 (2.9)	6.9 (1.1)	8.7 (2.7)	6.1 (1.0)
	101.4 (48.0)	41.2 (5.7)	91.0 (47.0)	52.9 (14.3)
$R_{p.i.m.}$* (%)	2.5 (0.9)	2.0 (0.3)	2.6 (0.9)	1.8 (0.3)
	28.8 (13.9)	11.6 (1.5)	26.8 (13.7)	15.5 (4.2)
$\langle \mathbf{I/\sigma(I)} \rangle$	22.9 (6.9)	26.5 (4.3)	21.6 (7.2)	29.3 (5.0)
	2.9 (1.1)	6.0 (1.4)	3.2 (1.7)	4.9 (1.4)
CC(1/2)•	0.9784 (0.0538)	0.9783 (0.0495)	0.9621 (0.0964)	0.9993 (0.0004)
	0.7930 (0.1432)	0.9163 (0.0828)	0.8143 (0.1321)	0.9434 (0.0251)

‡Maximum high resolution limit defined by AIMLESS using an I/σ(I) > 2 in the highest resolution shell. ▼Highest resolution bin based on I/σ(I) > 2. ⁺$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i (hkl)$ and *$R_{p.i.m.} = \sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i (hkl)$, where $I_i(hkl)$ represents observed intensity of the ith measurement of the reflection hkl and $\langle I(hkl)\rangle$ represent the mean intensity of reflection hkl after scaling. •CC(1/2) is calculated by $r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$ where x and y represent random half sets of reflection and n is the sample size.

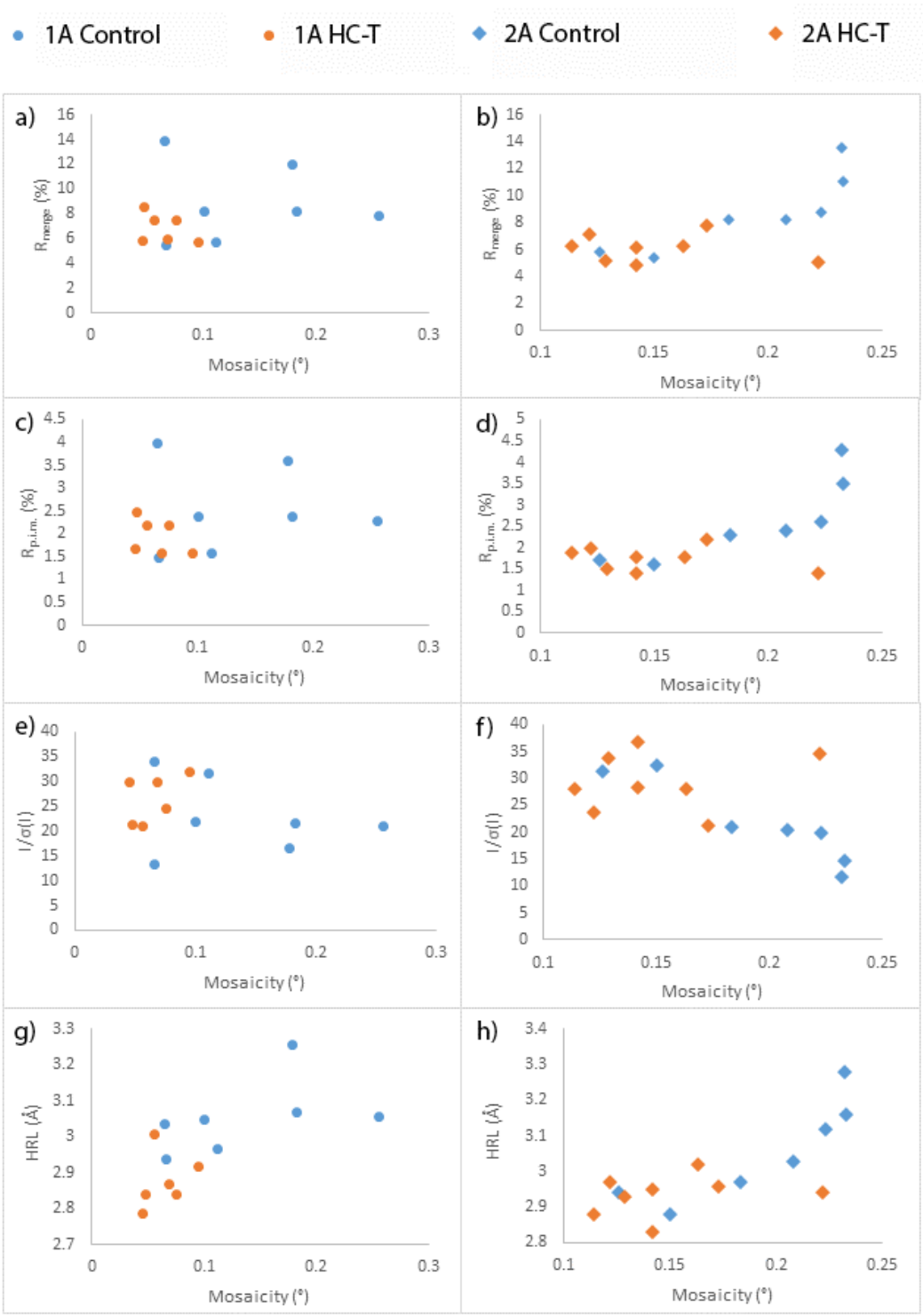


Fig. 3.7: Individual dataset overall data quality indicators plotted against mosaicity for bacteriocin syringacin M-1 crystal control (blue) and HC-transferred (orange) samples at λ = 1.0 Å (left/circles) and λ = 2.0 Å (right/diamonds). (a-b) Overall R_{merge} (%). (c-d) Overall $R_{p.i.m.}$ (%). (e-f) Overall mean I/σ(I). (g-h) Maximum high resolution limit (Å).

Table 3.3: Average data-collection statistics for bacteriocin syringacin M-two crystals at λ = 1 Å and λ = 1.8 Å. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below

Crystal	**Bacteriocin Syringacin M Two λ = 1 Å**		**Bacteriocin Syringacin M Two λ = 1.8 Å**	
Method	Control	HC-Transferred	Control	HC-Transferred
No. Crystals	8	10	10	10
Crystal Size (mm)	0.19 (0.02) x 0.17 (0.02) x 0.12 (0.01)	0.19 (0.02) x 0.17 (0.03) x 0.13 (0.01)	0.21 (0.04) x 0.18 (0.03) x 0.14 (0.02)	0.18 (0.02) x 0.17 (0.02) x 0.14 (0.02)
Relative Humidity (%)	N/A	89.5	N/A	89.5
Space Group	$P6_322$	$P6_322$	$P6_322$	$P6_322$
Unit-cell Parameters (Å)	a = b = 160.24 (0.32) c = 99.27 (0.29)	a = b = 160.36 (0.12) c = 99.76 (0.08)	a = b = 160.76 (0.59) c = 99.71 (0.28)	a = b = 160.56 (0.11) c = 99.75 (0.22)
Mosaicity (°)	0.19 (0.06)	0.10 (0.03)	0.14 (0.02)	0.13 (0.04)
Maximum High Resolution Limit (Å)•	3.37 (0.15)	3.11 (0.14)	3.17	3.06
High Resolution Bin Range (Å) •	3.41-3.74	3.41-3.74	3.17-3.39	3.17-3.39
Multiplicity	12.5 (0.4)	13.0 (0.3)	12.5 (0.5)	12.5 (1.8)
	13.2 (0.1)	13.4 (0.3)	12.4 (0.5)	12.3 (1.9)
Completeness (%)	99.3 (0.5)	98.8 (2.0)	93.8 (9.3)	93.9 (4.0)
	99.9 (0.2)	99.4 (1.5)	95.4 (8.1)	96.1 (2.9)
R_{merge} (%)•	11.7 (3.6)	6.6 (1.5)	7.5 (1.9)	5.7 (1.0)
	86.4 (23.2)	38.0 (18.6)	104.1 (53.9)	61.6 (18.8)
$R_{p.i.m.}$ (%)•	3.4 (1.2)	1.9 (0.4)	2.2 (0.6)	1.6 (0.1)
	23.2 (6.3)	10.5 (5.1)	30.2 (16.3)	17.9 (5.8)
$\langle \mathbf{I}/\boldsymbol{\sigma}(\mathbf{I}) \rangle$	18.5 (5.4)	32.0 (5.3)	24.8 (7.3)	31.4 (4.8)
	3.0 (1.1)	8.8 (4.4)	2.4 (1.0)	4.4 (1.9)
CC(1/2)•	0.9993 (0.0007)	0.9998 (0.0004)	0.9996 (0.0005)	0.9998 (0.0004)
	0.8741 (0.0587)	0.9618 (0.0316)	0.8256 (0.1232)	0.9212 (0.0481)

•See Table 3.2 footnote for definitions.

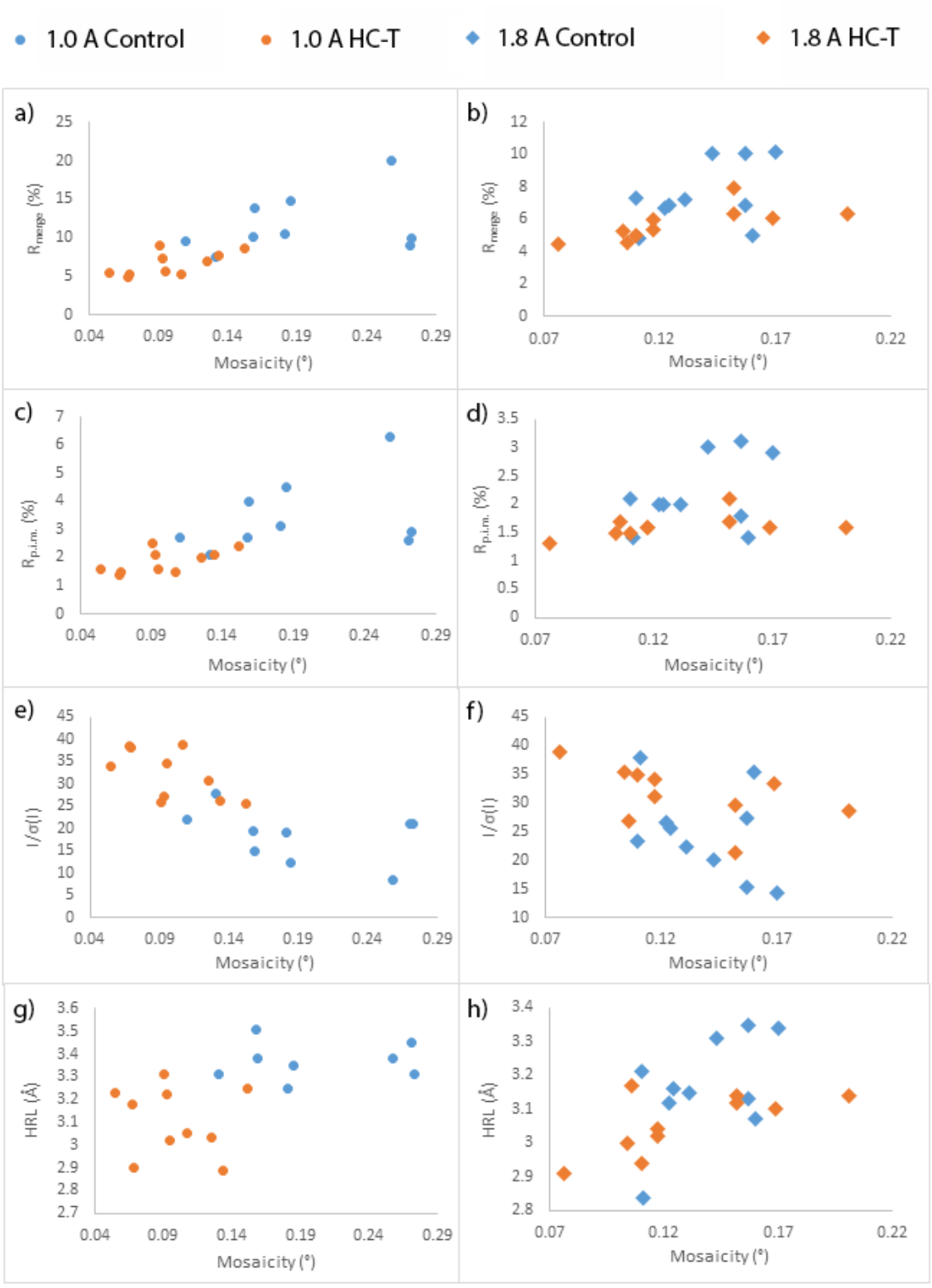


Fig. 3.8: Individual dataset overall data quality indicators plotted against mosaicity for bacteriocin syringacin M-2 crystal control (blue) and HC-transferred (orange) samples at λ = 1.0 Å (left/circles) and λ = 1.8 Å (right/diamonds). (a-b) Overall R_{merge} (%). (c-d) Overall $R_{p.i.m.}$ (%). (e-f) Overall mean I/σ(I). (g-h) Maximum high resolution limit (Å).

3.3.2.3. Bacteriocin syringacin M RADDOSE-3D Analysis

RADDOSE-3D analysis for bacteriocin syringacin M experiments predicts a slight increase in the elastic yield from 12.4 keV to 10.3 keV with a steady decrease thereafter with decreasing energy for both crystal and buffer (Fig. 3.9a). The diffraction efficiency shows a steady decrease with decreasing energy towards a plateau around 5.2 keV at 4.95E+11 photons/MGy. Buffer alone shows a marginal increase in diffraction efficiency over a crystal at higher energies, which gradually decreases with energy (Fig. 3.9b). Diffraction weighted dose increases to a peak around 0.06 MGy between 7.8 keV and 8.9 keV before decreasing back to 0.03 MGy at 5.2 keV (Fig. 3.9c). Diffraction weighted dose for buffer alone differs from the crystal at 10.3 keV, 7.8 keV and 6.2 keV with all 0.01 MGy lower (Fig. 3.9c). Maximum dose shows a similar trend to diffraction weighted dose, with an

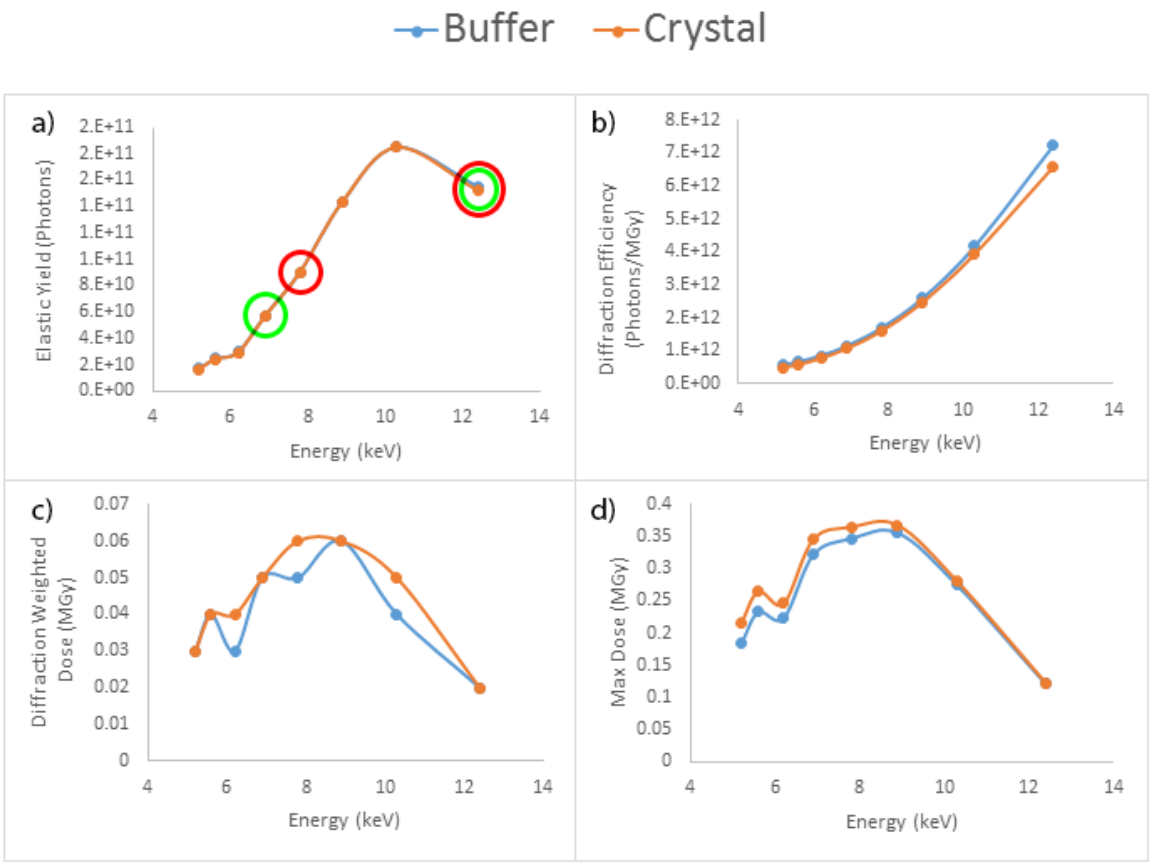


Fig. 3.9: RADDOSE-3D calculations for bacteriocin syringacin M crystals (orange) and buffer only (blue), between 12.4 keV and 5.2 keV. (a) Elastic yield (photons). (b) Diffraction efficiency (photons/MGy). (c) Diffraction weighted dose (MGy). (d) Maximum Dose (MGy). The energies used in data collection for BSM-1 were 12.7 keV (λ = 0.9795 Å) and 6.2 keV (λ = 2.0 Å), circled in green, and for BSM-2 were 12.4 keV (λ = 1.0 Å) and 6.9 keV (λ = 1.8 Å) circled in red.

initial increase to around 0.35 MGy between 6.9-8.9 keV which decreases at lower energies. Protein and buffer show little difference at higher energies but max dose decreases slightly for buffer alone at 8.9 keV and lower.

3.3.2.4. *PA*3825-EAL

Overall individual data quality indicators for control and HC-transferred *PA*3825-EAL datasets collected at λ = 1 Å and λ = 1.8 Å are presented in Figure 3.10, with outliers determined by the Grubb's test circled in red. Mosaicity and overall data quality indicators show near identical spread and correlation between treatments at λ = 1.0 Å and λ = 1.8 Å, although data collected at λ = 1.8 Å show a much larger distribution in mosaicity compared to λ = 1.0 Å data (Fig. 3.10). Both treatments at λ = 1.0 Å show outliers in mosaicity and/or data quality indicators, which were omitted from averages, standard deviations and statistical analysis.

Average values and standard deviations for *PA*3825-EAL for both treatments at λ = 1.0 Å and λ = 1.8 Å were calculated from 11 crystals from each population and are presented in Table 3.4. All average values show significant overlap between treatments and data points are well within one standard deviation (Fig. 3.14). This is reflected in the Student's t-test results where all data quality indicators, except λ = 1 Å overall I/σ(I), show no significant differences between treatments at λ = 1 Å or λ = 1.8 Å.

Table 3.4: Average data-collection statistics for *PA*3825-EAL at λ = 1 Å and λ = 1.8 Å. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below.

Test Crystal	***PA*3825-EAL λ = 1 Å**		***PA*3825-EAL λ = 1.8 Å**	
Method	Control	HC-Transferred	Control	HC-Transferred
No. Crystals	11	11	11	11
Crystal Size (mm)	0.23 (0.07) x 0.18 (0.06) x 0.16 (0.05)	0.22 (0.05) x 0.17 (0.04) x 0.16 (0.03)	0.18 (0.03) x 0.15 (0.03) x 0.13 (0.02)	0.17 (0.03) x 0.14 (0.02) x 0.12 (0.03)
Relative Humidity (%)	N/A	95	N/A	95
Space Group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell Parameters (Å)	a = b = 63.83 (0.21) c = 135.40 (0.42)	a = b = 64.03 (0.07) c = 135.70 (0.20)	a = b = 63.90 (0.15) c = 135.32 (0.49)	a = b = 64.04 (0.09) c = 135.64 (0.18)
Mosaicity (°)	0.08 (0.07)	0.076 (0.013)	0.12 (0.05)	0.15 (0.08)
Maximum High Resolution Limit (Å)•	1.86 (0.09)	1.87 (0.13)	1.86 (0.04)	1.85 (0.05)
High Resolution Bin Range (Å) •	1.88-1.92	1.88-1.92	1.86-1.90	1.86-1.90
Multiplicity	8.0 (0.1)	8.0 (0.2)	7.0 (0.2)	6.9 (0.1)
	5.1 (0.5)	5.1 (1.5)	2.7 (0.4)	2.3 (0.2)
Completeness (%)	99.1 (1.1)	99.6 (0.7)	95.5 (1.9)	97.7 (1.1)
	96.0 (6.7)	96.8 (6.5)	69.2 (8.5)	80.6 (6.1)
R_{merge}• (%)	6.9 (1.8)	5.5 (1.4)	4.9 (0.8)	4.24 (1.0)
	56.8 (17.9)	50.3 (24.1)	40.6 (9.1)	32.5 (11.1)
$R_{p.i.m.}$• (%)	2.5 (0.7)	2.0 (0.5)	1.9 (0.3)	1.6 (0.4)
	27.2 (8.8)	25.8 (14.9)	27.2 (6.2)	24.2 (8.6)
$\langle \mathbf{I}/\boldsymbol{\sigma}(\mathbf{I}) \rangle$	20.4 (7.2)	21.0 (3.3)	24.3 (4.2)	26.5 (7.7)
	3.0 (2.6)	2.5 (1.1)	2.3 (0.5)	2.7 (1.4)
CC(1/2) •	0.9990 (0.0000)	0.9991 (0.0013)	0.9994 (0.0005)	0.9997 (0.0004)
	0.8207 (0.1048)	0.8038 (0.2453)	0.6036 (0.1802)	0.6232 (0.3073)

•See Table 3.2 footnote for definitions.

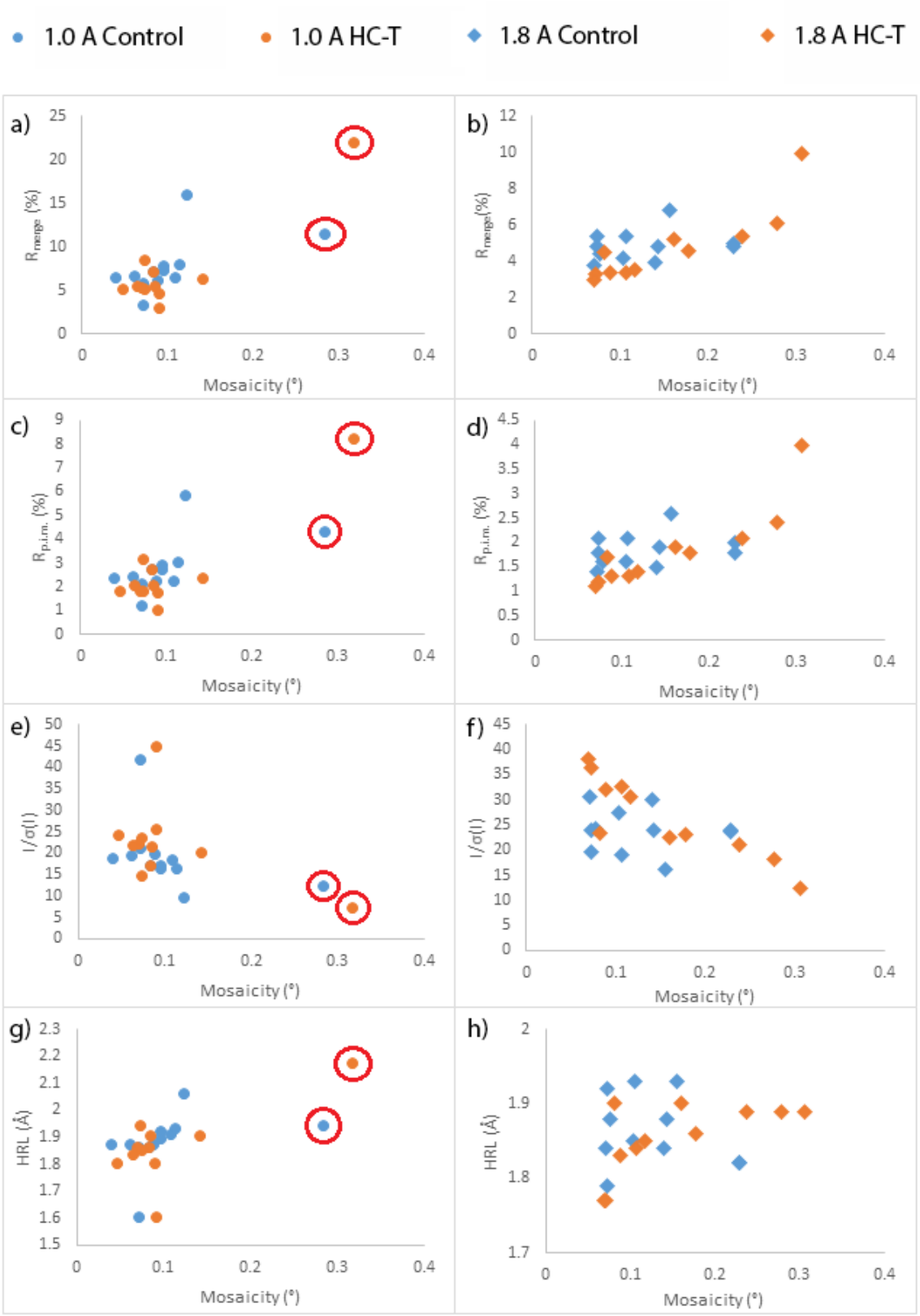


Fig. 3.10: Individual dataset overall data quality indicators plotted against mosaicity for *PA*3825-EAL crystal control (blue) and HC-transferred (orange) samples at λ = 1.0 Å (left/circles) and λ = 1.8 Å (right/diamonds). (a-b) Overall R_{merge} (%). (c-d) Overall $R_{p.i.m.}$ (%). (e-f) Overall mean I/σ(I). (g-h) Maximum high resolution limit (Å).

3.3.2.5. *PA*3825-EAL RADDOSE-3D Analysis

Analysis with RADDOSE-3D shows an initial dip in the elastic yield and diffraction efficiency from 12.4 keV to 10.3 keV, followed by a significant increase at 8.9 Kev and a gradual decrease below 8.9 keV (Fig. 3.11). Values for the buffer show no variation from the crystal in elastic yield (Fig. 3.11a) but an increased diffraction efficiency over all energies (Fig. 3.11b). Diffraction weighted dose and maximum dose increase between 12.4 keV and 10.3 keV, followed by a gradual decrease at lower energies. Buffer alone has a consistently lower diffraction weighted dose and maximum dose compare to the crystal at all energies (Fig. 3.11c-d), accounting for the higher diffraction efficiency observed in Fig. 3.11b. RADDOSE-3D calculations suggest there is a slight increase in the elastic yield and diffraction efficiency of crystals collected at λ = 1.8 Å compared to λ = 1.0 Å, with a 1.3 and 1.9 times increase, respectively (Fig. 3.11a-b). Diffraction weighted dose and maximum dose both

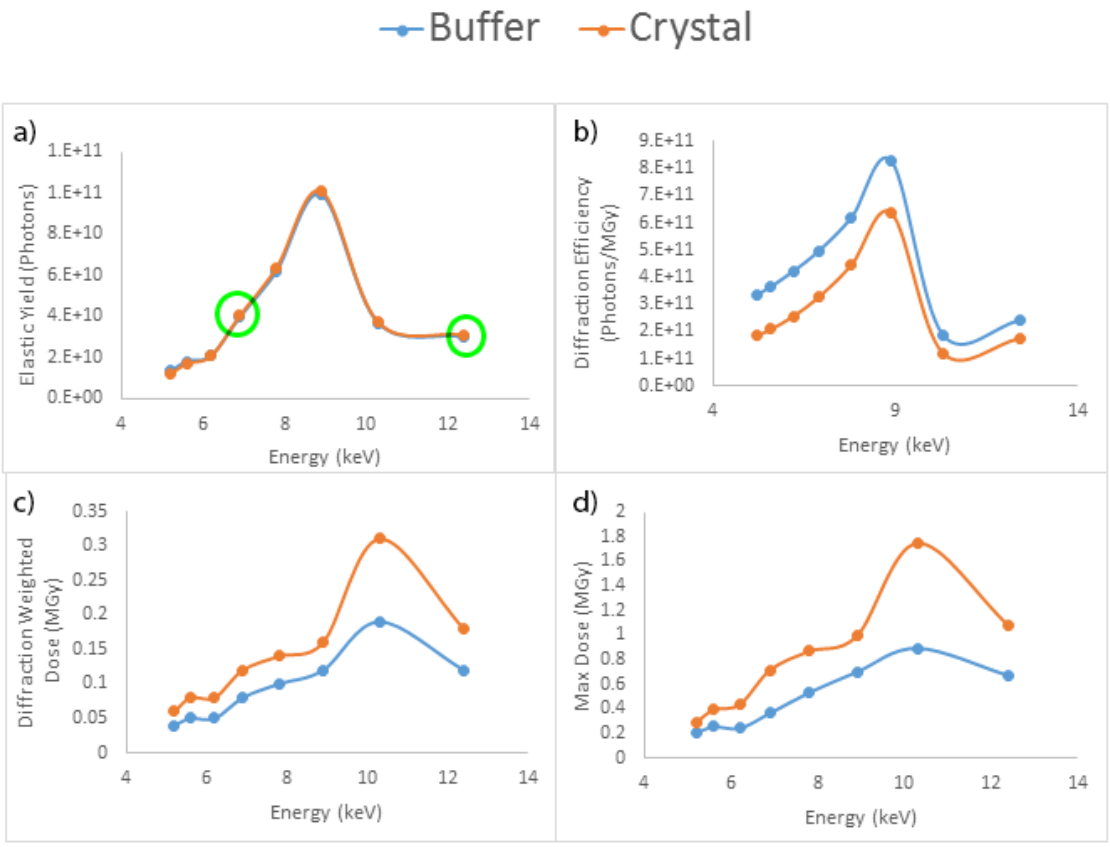


Fig. 3.11: RADDOSE-3D calculations for PA3825-EAL crystals (orange) and buffer only (blue), between 12.4 keV and 5.2 keV. (a) Elastic yield (photons). (b) Diffraction efficiency (photons/MGy). (c) Diffraction weighted dose (MGy). (d) Maximum dose (MGy). The energies used in data collection for PA3825-EAL were 12.4 keV (λ = 1.0 Å) and 6.9 keV (λ = 1.8 Å) circled in green.

decrease at λ = 1.8 Å compared to λ = 1.0 Å with a 1.5 times decrease in both dose calculations (Fig. 3.11c-d).

3.3.2.6. Bacteriocin Pectocin M2

Individual data quality indicators plotted against mosaicity for BMP2 present a clear trend between treatments collected at both wavelengths. Control datasets show a pronounced skew towards improvement in mosaicity, R_{merge}, $R_{p.i.m.}$, I/σ(I) and high resolution limit when compared to HC-transferred samples (Fig. 3.12). This is particularly evident in λ = 1 Å datasets where control mosaicity is clustered between 0.048° and 0.082° (excluding Grubb's test outliers circled in red in figure 3.12) and HC-transferred datasets spread between 0.054° and 0.592° (Fig. 3.12a,c, e and g). The difference between data quality indicators and mosaicity between treatments is less pronounced in λ = 2 Å datasets, but still suggest an improvement over HC-transferred samples (Fig. 8.11b, d, f and h).

Average data quality indicators and standard deviations for control/HC-transferred populations were calculated from 16/11 and 12/10 crystals for λ = 1 Å and λ = 2 Å, respectively, and are presented in Table 3.5. Statistical analysis shows a significant improvement in data quality indicators of control samples at λ = 1 Å and λ = 2 Å, with the exception of overall R_{merge} and $R_{p.i.m.}$ at λ = 2 Å which show no significant difference. These differences are evident in average data quality indicators plotted in figure 3.14, where data points are over or close to one standard deviation apart in the vertical direction, with the exception of overall $R_{p.i.m.}$ collected at λ = 2 Å (Fig. 3.14a). High resolution limits for BPM2 at λ = 2 Å are within one standard deviation (Fig. 3.14), but are still significantly different (Table 3.6).

Table 3.5. HC1 Vs Control: Average data-collection statistics for bacteriocin pectocin M2 crystals at λ = 1 Å and λ = 1.8 Å. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below.

Crystal	**Bacteriocin Pectocin M2 λ = 1.0 Å**		**Bacteriocin Pectocin M2 λ = 2.0 Å**	
Method	Control	HC-Transferred	Control	HC-Transferred
No. Crystals	16	11	12	10
Crystal Size (mm)	0.15 (0.02) x 0.10 (0.02) x 0.09 (0.01)	0.13 (0.03) x 0.08 (0.02) x 0.09 (0.02)	0.16 (0.02) x 0.11 (0.02) x 0.09 (0.02)	0.15 (0.03) x 0.10 (0.02) x 0.09 (0.01)
Relative Humidity (%)	N/A	80.5	N/A	80.5
Space Group	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Unit-cell Parameters (Å)	a = b = 117.30 (0.39) c = 128.42 (0.31)	a = b = 116.39 (0.64) c = 128.12 (0.34)	a = b = 117.49 (0.09) c = 128.39 (0.15)	a = b = 117.02 (0.53) c = 128.59 (0.41)
Mosaicity (°)	0.062 (0.010)	0.227 (0.147)	0.121 (0.029)	0.184 (0.040)
Maximum High Resolution Limit (Å)•	2.16 (0.17)	2.64 (0.4)	2.28 (0.05)	2.46 (0.1)
High Resolution Bin Range (Å) •	2.63-2.76	2.63-2.76	2.61-2.73	2.61-2.73
Multiplicity	10.0 (0.1)	9.5 (0.6)	9.6 (0.2)	9.4 (0.1)
	10.0 (0.1)	9.2 (1.1)	9.6 (0.4)	9.5 (0.2)
Completeness (%)	99.8 (0.0)	94.6 (8.5)	99.0 (1.4)	99.2 (1.2)
	100.0 (0.0)	90.0 (16.0)	98.4 (2.7)	98.5 (2.5)
R_{merge}• (%)	5.7 (3.4)	8.3 (2.8)	9.0 (2.0)	9.7 (1.5)
	24.6 (25.9)	77.2 (57.4)	39.9 (13.7)	66.4 (18.7)
$R_{p.i.m.}$• (%)	1.9 (1.1)	2.9 (1.0)	3.1 (0.7)	3.3 (0.5)
	8.2 (8.6)	27.7 (22.0)	13.5 (4.7)	22.6 (6.3)
$\langle \mathbf{I}/\boldsymbol{\sigma}(\mathbf{I}) \rangle$	34.0 (9.0)	21.0 (7.9)	21.1 (3.9)	17.2 (2.5)
	12.7 (4.6)	5.2 (4.4)	7.9 (2.0)	4.6 (1.5)
CC(1/2)•	0.9993 (0.0006)	0.9991 (0.0005)	0.9976 (0.0009)	0.9985 (0.0005)
	0.9799 (0.0423)	0.8155 (0.2309)	0.9741 (0.0227)	0.9311 (0.0426)

•See Table 3.2 footnote for definitions.

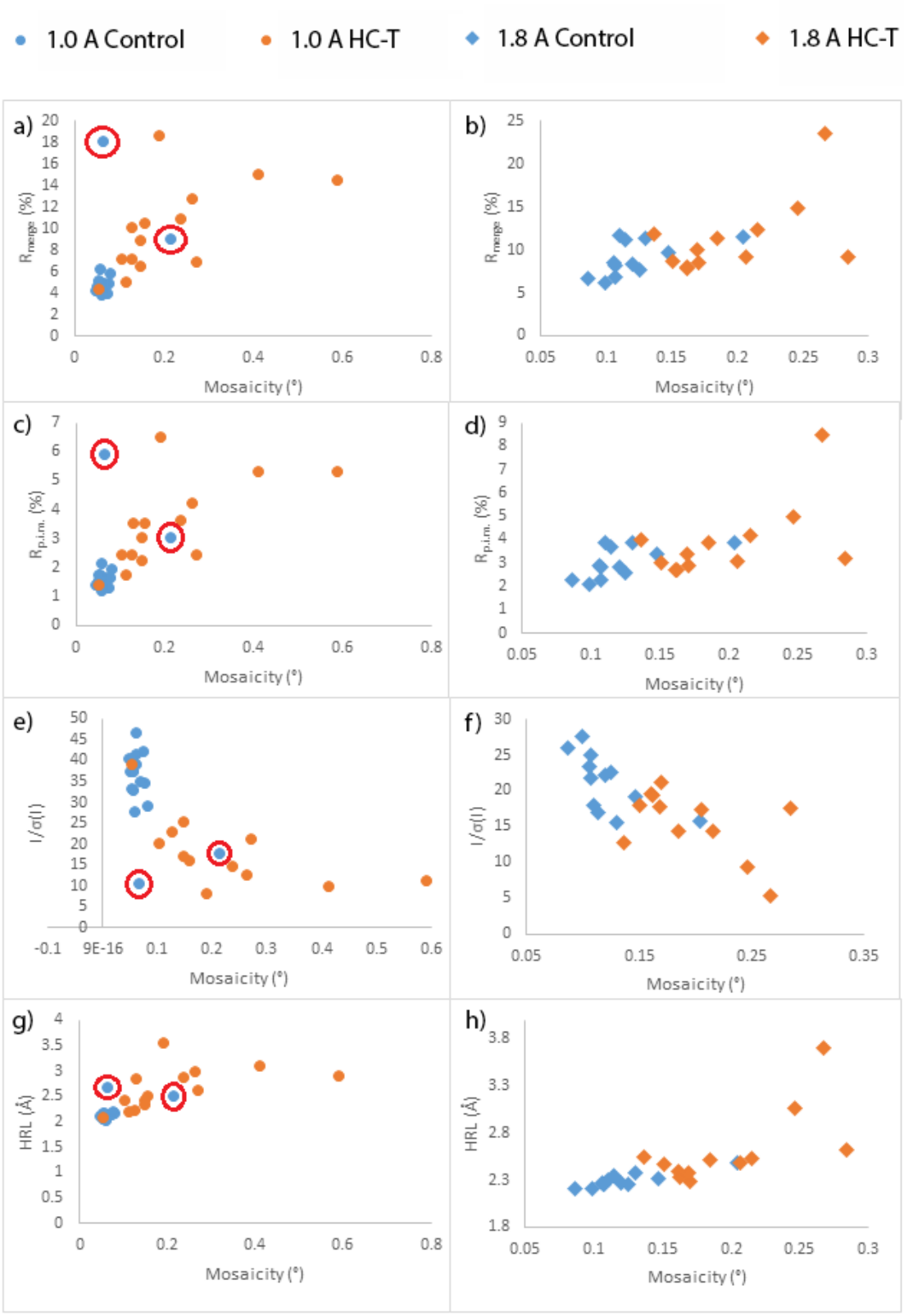


Fig. 3.12: Individual dataset overall data quality indicators plotted against mosaicity for bacteriocin pectocin M2 crystal control (blue) and HC-transferred (orange) samples at λ = 1.0 Å (left/circles) and λ = 1.8 Å (right/diamonds). (a-b) Overall R_{merge} (%). (c-d) Overall $R_{p.i.m.}$ (%). (e-f) Overall mean I/σ(I). (g-h) Maximum high resolution limit (Å).

3.3.2.7. Bacteriocin Pectocin M2 RADDOSE-3D Analysis

As with the BSM crystal, BPM2 sees a slight increase in elastic yield at 10.3 keV followed by a steady decrease with decreasing energy (Fig. 3.13a); as well as a steady decrease in diffraction efficiency with decreasing energy (Fig. 3.13b). Diffraction weighted dose and maximum dose also repeat the pattern seen in BSM with an increase to a maximum around 6.9-7.8 keV and decreasing thereafter (Fig. 3.13c-d). Elastic yield shows no difference between crystal orientations or the buffer alone (Fig. 3.13a). Crystal orientation two, with the beam passing through the long z-axis of the crystal, shows a 1.4 times increase in diffraction efficiency at 12.4 keV compared to crystal orientation one. This gradually decreases with the energy until there is little to no difference around

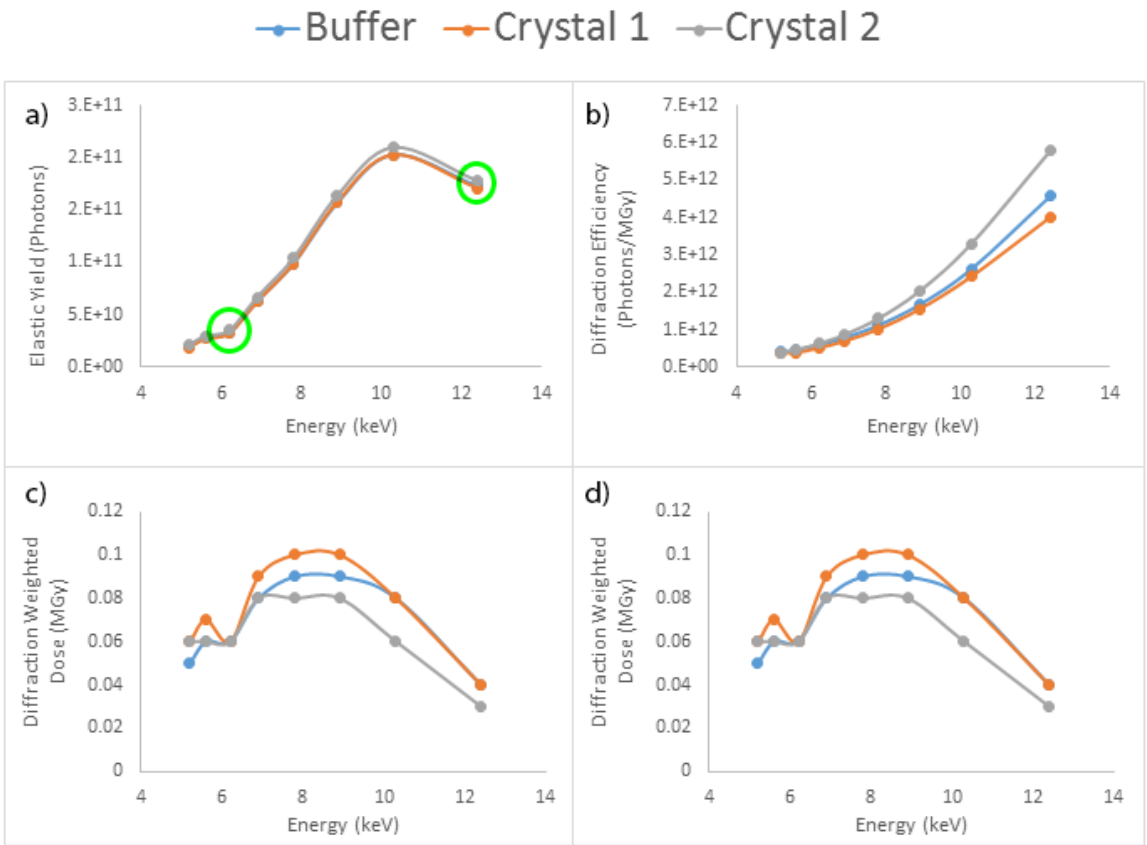


Fig. 3.13: RADDOSE-3D calculations for bacteriocin pectocin M2 crystals 1 (orange), 2 (grey) and buffer only (blue), between 12.4 keV and 5.2 keV. (a) Elastic yield (photons). (b) Diffraction efficiency (photons/MGy). (c) Diffraction weighted dose (MGy). (d) Maximum dose (MGy). The energies used in data collection for bacteriocin pectocin M2 were 12.4 keV (λ = 1.0 Å) and 6.2 keV (λ = 2.0 Å) circled in green.

7.8 keV (Fig. 3.13c). Buffer using the same dimensions as crystal orientation one shows a very similar curve to that of crystal orientation one. A Similar pattern is seen in diffraction weighted dose and maximum dose, with crystal orientation two presenting with a lower diffraction weighted dose and maximum dose at higher energies compared to crystal orientation one and buffer, until they become equal around 7.8 keV (Fig. 3.13c-d).

BPM2 crystals show a higher elastic yield and diffraction efficiency at higher energies, with a 5.3 and 7.8 times reduction, respectively, between λ = 1.0 Å and λ = 2.0 Å (Fig. 3.13a-b). Data collections at λ = 1.0 Å and λ = 2.0 Å are above and below the absorption edge of iron (λ = 1.74 Å), respectively, so absorption is not as severe as data collection at the iron edge. However, diffraction weighted dose and maximum dose is still increased by 0.020 MGy and 0.182 MGy, respectively, at λ = 2.0 Å compared to data collection at λ = 1.0 Å (Fig. 3.11c-d).

Table 3.6: Student's t-test results comparing control and HC-transferred populations for all data quality indicators. + represents populations that are statistically the same and - represents populations that are statistically different. Symbols in parentheses are for data in the highest resolution bin ranges defined in Table 3.2, 3.3, 3.4 and 3.5.

Crystal	**Bacteriocin Syringacin M One**		**Bacteriocin Syringacin M Two**		***PA*3825-EAL**		**Bacteriocin Pectocin M2**	
Wavelength (Å)	1.0	2.0	1.0	1.8	1.0	1.8	1.0	2.0
Unit Cell Parameter (a=b=) (Å)	+	+	+	+	-	-	-	-
Unit Cell Parameter (c=) (Å)	-	-	-	+	-	+	+	+
Mosaicity (°)	-	-	-	+	+	+	-	-
R_{merge}• (%)	+(-)	-(-)	-(-)	-(-)	+ (+)	+ (+)	-(-)	+ (-)
$R_{p.i.m.}$ • (%)	+(-)	-(-)	-(-)	-(-)	+ (+)	+ (+)	-(-)	+ (-)
I/σ(I)	+(-)	-(+)	-(-)	-(-)	- (+)	+ (+)	-(-)	-(-)
Maximum High Resolution Limit (Å) •	-	-	-	+	+	+	-	-

• See Table 3.2 footnote for definitions

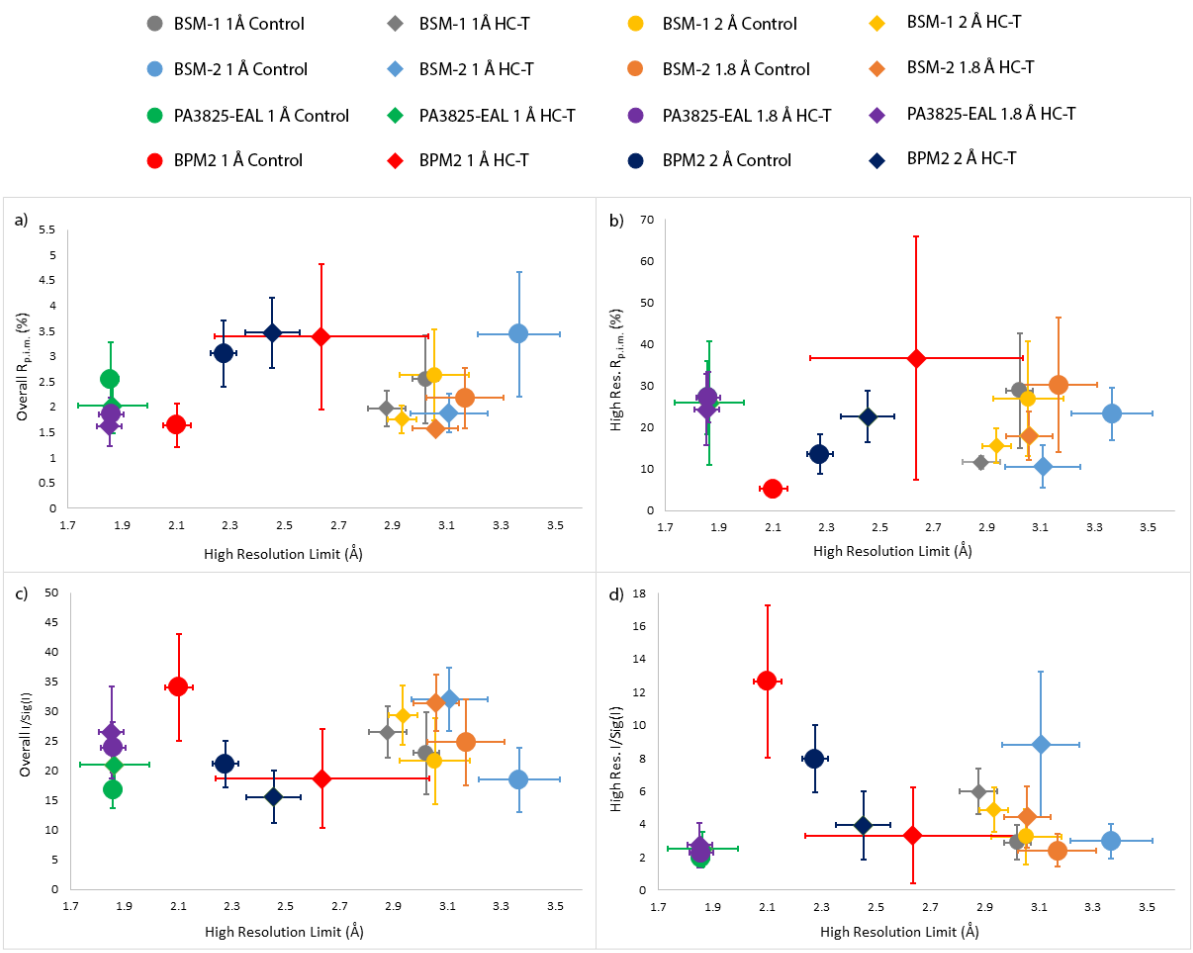


Fig. 3.14: Bacteriocin syringacin M-1 and -2, PA3825-EAL and bacteriocin pectocin M2 average data quality indicators for control (circles) and HC-transferred (diamonds) populations collected at λ = 1 Å and λ ≈ 2 Å plotted against average maximum high resolution limit. Error bars represent one standard deviation. (a) Overall $R_{p.i.m.}$. (b) High resolution $R_{p.i.m.}$. (c) Overall mean I/σ(I). (c) High resolution mean I/σ(I).

Like Vs Like Treatments at Different Wavelengths

Bacteriocin Syringacin M-one

BSM-one populations showed a significant improvement in data quality indicators following exposure to the humid air stream with a decrease in the mosaicity of HC-transferred samples and some differences in unit cell parameters. To further investigate this, control samples collected at λ = 1 Å and λ = 2 Å and HC-transferred samples collected at λ = 1 Å and λ = 2 Å were compared using the same statistical tests described previously. Average data quality indicators were

calculated from 7/7 and 7/8 λ = 1 Å/ λ = 2 Å crystals for control and HC-transferred populations, respectively, and are presented in Table 3. 7.

Comparison of the control crystals at different wavelengths showed consistent unit cell parameters and mosaicity (Table 3.10), this is evident from the similar spread of data points along the x-axis between control samples collected at λ = 1.0 Å (orange circles) and λ = 2.0 Å (blue circles) in figure 3.15a, c, e and g. HC-transferred samples collected at λ = 1.0 Å and λ = 2.0 Å also show consistent unit cell parameters, however, mosaicity is significantly improved in HC-transferred populations collected at λ = 1 Å. A clear divide in mosaicity is seen between λ = 1 Å and λ = 2 Å HC-transferred populations in figure 3.15b, d, f and h, with all λ = 1 Å data points (orange circles) being lower than the lowest λ = 2 Å data point (blue circles).

Comparison of average data quality indicators for BSM-1 control populations collected at λ = 1 Å and λ = 2 Å and for the HC-transferred populations at λ = 1 Å and λ = 2 Å showed no significant differences (Table 3.10). The similarity in data quality indicators is evident in figure 3.18, where control populations at both wavelengths (grey circles) are within one standard deviation, as are HC-transferred populations (yellow diamonds).

Table 3.7: Control vs Control and HC1 vs HC1 average data-collection statistics for bacteriocin syringacin M-one at λ = 1 Å and λ = 2.0 Å. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below.

Crystal	Bacteriocin Syringacin M-1 Control		Bacteriocin Syringacin M-1 HC-transferred	
Wavelength (Å)	1.0	2.0	1.0	2.0
No. Crystals	7	7	7	8
Crystal Size (mm)				
Relative Humidity (%)	N/A	N/A	89.0	89.0
Space Group	$P6_322$	$P6_322$	$P6_322$	$P6_322$
Unit-cell Parameters (Å)	a = b = 161.09 (0.50) c = 99.26 (0.35)	a = b = 161.1 (0.20) c = 99.37 (0.11)	a = b = 160.76 (0.47) c = 99.73 (0.01)	a = b = 161.00 (0.31) c = 99.76 (0.26)
Mosaicity (°)	0.137 (0.065)	0.194 (0.039)	0.064 (0.017)	0.151 (0.033)
Maximum High Resolution Limit (Å)•	3.02 (0.05)	3.05 (0.13)	2.88 (0.07)	2.94 (0.05)
High Resolution Bin Range (Å) •	3.26-3.05	3.26-3.05	3.26-3.05	3.26-3.05
Multiplicity	12.2 (1.8)	11.5 (1.8)	12.1 (2.6)	12.5 (0.2)
	12.3 (1.8)	11.6 (1.8)	12.2 (2.7)	12.3 (0.1)
Completeness (%)	99.7 (0.3)	99.5 (0.2)	98.6 (1.8)	98.9 (1.4)
	100.0 (0.0)	99.9 (0.2)	99.2 (1.4)	99.7 (0.6)
R_{merge}• (%)	8.8 (2.9)	8.7 (2.7)	6.9 (1.1)	6.1 (1.0)
	101.4 (48.0)	91.0 (47.0)	41.2 (5.7)	52.9 (14.3)
$R_{p.i.m.}$• (%)	2.5 (0.9)	2.6 (0.9)	2.0 (0.3)	1.8 (0.3)
	28.8 (13.9)	26.8 (13.7)	11.6 (1.5)	15.5 (4.2)
$\langle \mathbf{I}/\boldsymbol{\sigma}(\mathbf{I}) \rangle$	22.9 (6.9)	21.6 (7.2)	26.5 (4.3)	29.3 (5.0)
	2.9 (1.1)	3.2 (1.7)	6.0 (1.4)	4.9 (1.4)
CC(1/2)•	0.9784 (0.0538)	0.9621 (0.0964)	0.9783 (0.0495)	0.9993 (0.0004)
	0.7930 (0.1432)	0.8143 (0.1321)	0.9163 (0.0828)	0.9434 (0.0251)

• See Table 3.2 footnote for definitions.

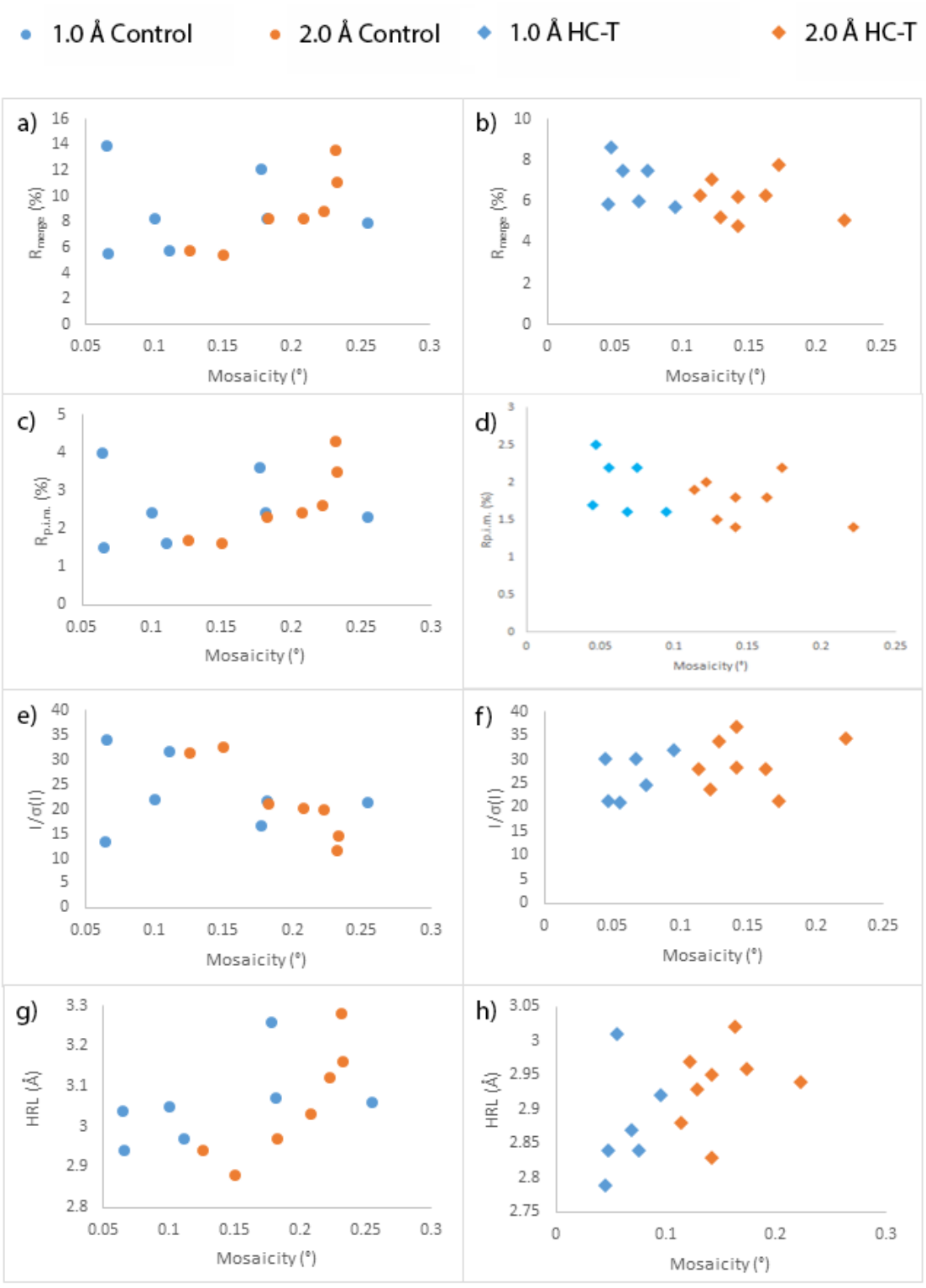


Fig. 3.15: Individual dataset overall data quality indicators plotted against mosaicity for like Vs like treatments for bacteriocin syringacin M-1 crystals. Control Vs control crystals (left, circles) and HC-transferred vs HC-trasnferred crystals (right, diamonds) are shown for datasets collected at λ = 1.0 Å (blue) and λ = 2.0 Å (orange). (a-b) Overall R_{merge} (%). (c-d) Overall $R_{p.i.m.}$ (%). (e-f) Overall mean I/σ(I). (g-h) Maximum high resolution limit (Å).

Bacteriocin Syringacin M-two

As with BSM-one, BSM-two crystals showed a significant improvement in data quality following transfer in the humid air stream with reproducible increases in high resolution limit and other data quality indicators. Average data quality indicators and standard deviations were calculated from λ = 1 Å/λ = 1.8 Å data collections from 8/10 and 10/10 crystals, respectively, for control and HC-transferred populations; and are presented in Table 3.8.

Control populations collected at λ = 1 Å show a much larger variance in mosaicity compared to the control population collected at λ =1.8 Å, with no outliers flagged by the Grubbs test (Fig. 3.16a, c, e and g). This is reflected in the average values, with a statistically significant increase in mosaicity of 0.05° in λ = 1 Å controls (Table 3.8). λ =1.8 Å controls show much smaller variation in mosaicity, with data points laying between a minimum of 0.11° and maximum of 0.17°. In addition to the increase in mosaicity, λ = 1 Å controls show a significant decrease in all unit cell parameters (Table 3.10). HC-transferred BSM-two crystals are much more stable in the humid air stream, with both populations collected at λ = 1 Å and λ = 1.8 Å showing a similar spread in mosaicity (Fig. 3.16b, d, f and h) and no significant differences by the t-test (Table 3.10). Unit cell parameters show a similar significant increase in a and b unit cell parameters in the λ = 1.8 Å population but the c axis remains unchanged between populations (Table 3.10).

Control populations show a consistent improvement in overall and high resolution R_{merge}, $R_{p.i.m.}$ and I/σ(I) at λ = 1.8 Å (Table 3.10) while HC-transferred populations only see an improvement in overall R_{merge}, $R_{p.i.m.}$ and I/σ(I) at λ = 1.8 Å (Table 3.10). This is clearly observed in figure 3.18, where control BSM two data at λ = 1.8 Å (filled blue circles) have much lower average $R_{p.i.m.}$ and higher I/σ(I) compared to data at λ = 1.0 Å (open blue circles). A similar trend is observed in high resolution limit, with λ = 1.8 Å control data showing a significant improvement over λ = 1.0 Å control data. The HC-transferred populations show no significant differences in the high resolution limit at

the two wavelengths, and their average high resolution $R_{p.i.m.}$ and I/σ(I) show no significant differences (Table 3.10).

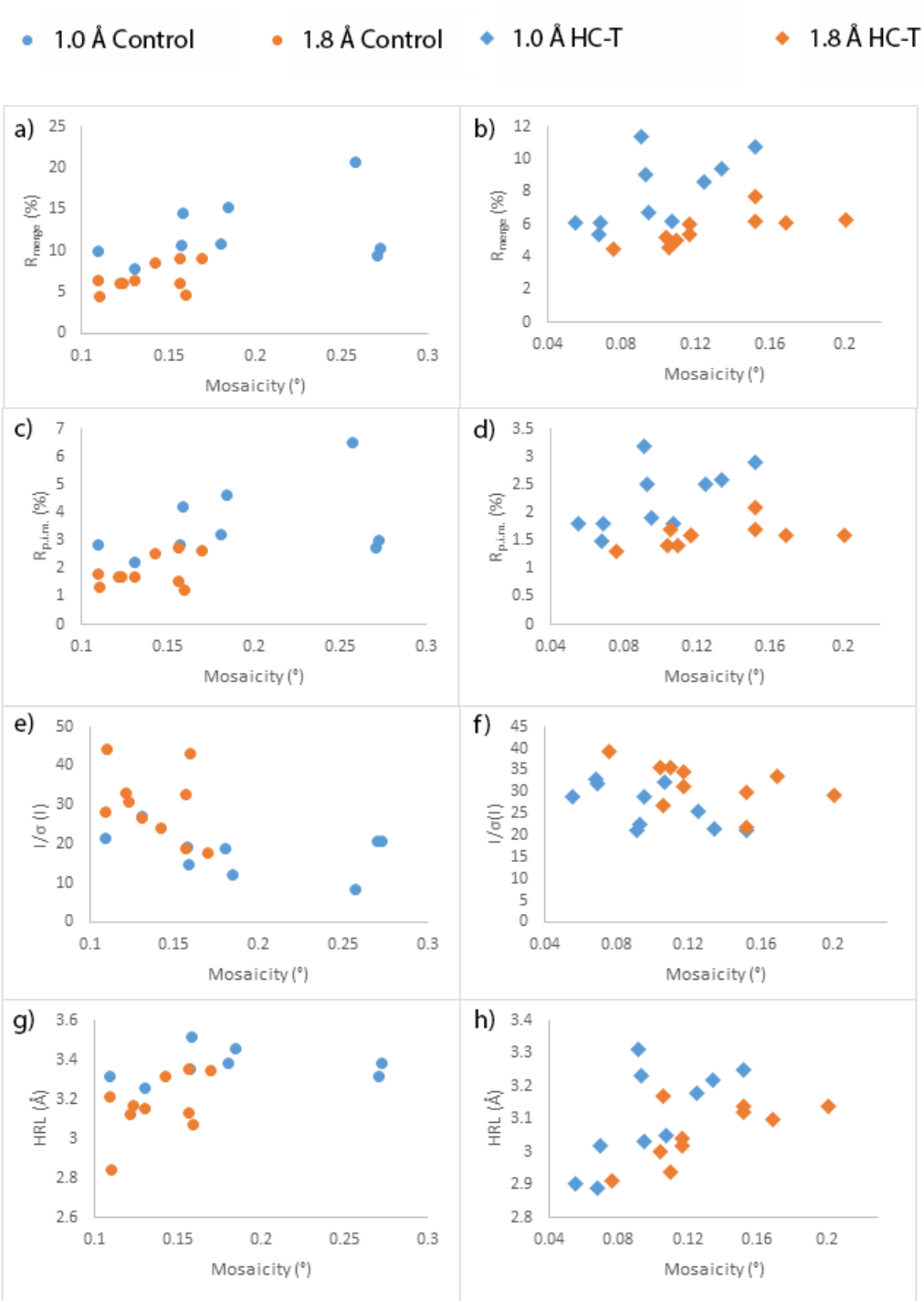


Fig. 3.16: Individual dataset overall data quality indicators plotted against mosaicity for like Vs like treatments for bacteriocin syringacin M-2 crystals. Control Vs control crystals (left, circles) and HC-transferred vs HC-trasnferred crystals (right, diamonds) are shown for datasets collected at λ = 1.0 Å (blue) and λ = 1.8 Å (orange). (a-b) Overall R_{merge} (%). (c-d) Overall $R_{p.i.m.}$ (%). (e-f) Overall mean I/σ(I). (g-h) Maximum high resolution limit (Å).

Table 3.8: Control vs Control and HC1 vs HC1 average data-collection statistics for bacteriocin syringacin M-two at λ = 1 Å and λ = 1.8 Å. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I /σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below.

Crystal	Bacteriocin Syringacin M-2 Control		Bacteriocin Syringacin M-2 HC-transferred	
Wavelength (Å)	1.0	1.8	1.0	1.8
No. Crystals	8	10	10	10
Crystal Size (mm)				
Relative Humidity (%)	N/A	N/A	89.5	89.5
Space Group	$P6_322$	$P6_322$	$P6_322$	$P6_322$
Unit-cell Parameters (Å)	a = b = 160.24 (0.32) c = 99.27 (0.29)	a = b = 160.76 (0.59) c = 99.71 (0.28)	a = b = 160.36 (0.12) c = 99.76 (0.08)	a = b = 160.56 (0.11) c = 99.75 (0.22)
Mosaicity (°)	0.19 (0.06)	0.14 (0.02)	0.10 (0.03)	0.13 (0.04)
Maximum High Resolution Limit (Å)•	3.37 (0.15)	3.17 (0.14)	3.11 (0.14)	3.06
High Resolution Bin Range (Å) •	3.64-3.37	3.64-3.37	3.39-3.17	3.39-3.17
Multiplicity	12.4 (0.5)	12.7 (0.5)	13.1 (0.3)	12.7 (1.8)
	13.0 (0.1)	13.2 (0.5)	13.4 (0.3)	12.4 (1.9)
Completeness (%)	99.3 (0.4)	93.8 (9.6)	98.9 (1.9)	94.0 (4.1)
	99.9 (0.2)	95.4 (8.4)	99.5 (1.6)	96.0 (3.0)
R_{merge}• (%)	12.1 (3.8)	6.6 (1.6)	7.8 (2.0)	5.7 (0.9)
	100.6 (26.6)	43.5 (20.5)	94.5 (53.4)	61.6 (18.74)
$R_{p.i.m.}$• (%)	3.6 (1.3)	1.9 (0.5)	2.3 (0.5)	1.6 (0.2)
	28.3 (7.6)	12.3 (6.1)	26.2 (14.6)	17.8 (5.8)
$\langle \mathbf{I}/\boldsymbol{\sigma}(\mathbf{I}) \rangle$	18.0 (5.3)	29.7 (8.4)	26.7 (4.7)	31.8 (4.8)
	2.4 (0.9)	7.4 (4.2)	4.0 (2.3)	4.6 (1.9)
CC(1/2)•	0.9992 (0.0009)	0.9996 (0.0005)	0.9998 (0.0004)	0.9999 (0.0003)
	0.7887 (0.1345)	0.9536 (0.0351)	0.8550 (0.1134)	0.9227 (0.0493)

• See Table 3.2 footnote for definitions

3.3.3.3. Bacteriocin Pectocin M2

Although exposure of BPM2 crystals to the humid air stream shows a detrimental effect on data quality, particularly in the high resolution limit. The effect is reproducible and the HC-transferred and control populations at λ = 1 Å and λ = 2 Å can be analysed for differences in data quality as above. Average data quality indicators and standard deviations were calculated from λ = 1 Å/λ = 2 Å data collections from 16/11 and 12/10 crystals, respectively, for control and HC-transferred populations; and are presented in Table 3.9.

Individual data quality indicators are plotted against mosaicity in figure 3.17 for control populations and for HC-transferred populations at both wavelengths with Grubb's test outliers circled in red. Control populations collected at λ = 1 Å show lower mosaicity and improved data quality indicators compared to control data collected at λ = 2 Å (Fig. 3.17a, c, e and g). However, the same trend is not observed for HC-transferred data, which show comparable spreads in mosaicity and all data quality indicators at both wavelengths (Fig. 3.17 b, d, f and h).

Control populations show a significant improvement in data quality at λ = 1 Å (Table 3.10). Overall and high resolution R_{merge}, $R_{p.i.m.}$, I/σ(I) and high resolution limit all show significant improvements over average data collected at λ = 2 Å. This is evident in figure 3.18 where control populations (blue) are well over one standard deviation apart. The same effect is not observed between HC-transferred crystals, with all overall and high resolution data quality indicators all within one standard deviation of one another (Fig. 3.18) and showing no significant difference between wavelengths (Table 3.10).

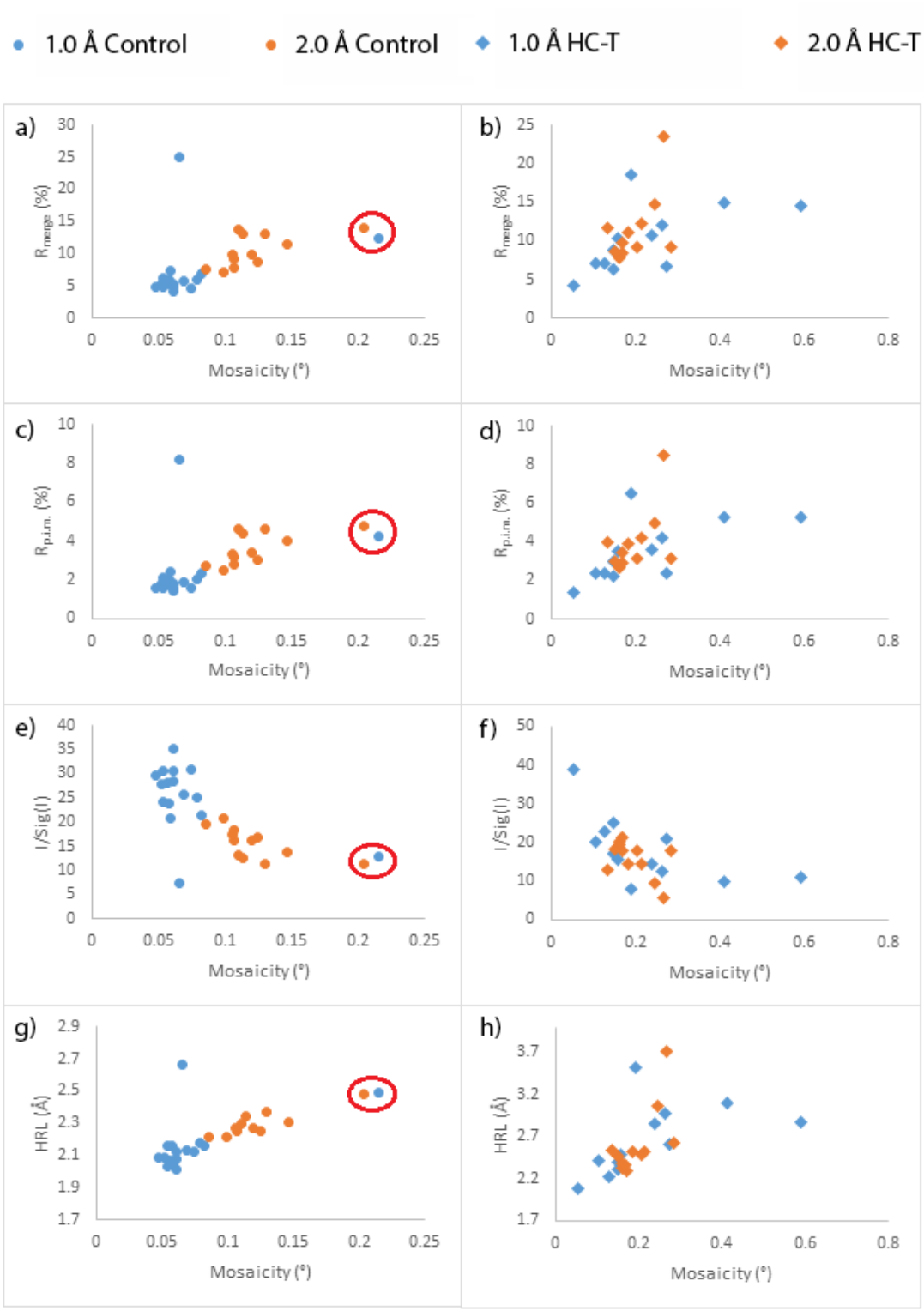


Fig. 3.17: Individual dataset overall data quality indicators plotted against mosaicity for like Vs like treatments for bacteriocin pectocin M2 crystals. Control Vs control crystals (left, circles) and HC-transferred vs HC-trasnferred crystals (right, diamonds) are shown for datasets collected at λ = 1.0 Å (blue) and λ = 2.0 Å (orange). (a-b) Overall R_{merge} (%). (c-d) Overall $R_{p.i.m.}$ (%). (e-f) Overall mean I/σ(I). (g-h) Maximum high resolution limit (Å).

Table 3.9: Control vs Control and HC1 vs HC1 average data-collection statistics for bacteriocin Pectocin M2 at λ = 1 Å and λ = 1.8 Å. Values in parentheses show one standard deviation. Values for multiplicity, completeness, R_{merge}, $R_{p.i.m.}$, mean I/σ(I) and CC(1/2) are shown as overall statistics followed by the highest resolution bin below.

Crystal	**Bacteriocin Pectocin M2 Control**		**Bacteriocin Pectocin M2 HC-transferred**	
Wavelength (Å)	1.0	2.0	1.0	2.0
No. Crystals	15	12	14	12
Crystal Size (mm)				
Relative Humidity (%)				
Space Group	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Unit-cell Parameters (Å)	a = b = 117.30 (0.39) c = 128.42 (0.31)	a = b = 117.49 (0.09) c = 128.39 (0.15)	a = b = 116.39 (0.64) c = 128.12 (0.34)	a = b = 117.02 (0.53) c = 128.59 (0.41)
Mosaicity (°)	0.07 (0.04)	0.121 (0.029)	0.227 (0.147)	0.184 (0.040)
Maximum High Resolution Limit (Å)•	2.16 (0.17)	2.28 (0.05)	2.64 (0.4)	2.46 (0.1)
High Resolution Bin Range (Å) •	2.29-2.37	2.29-2.37	2.63-2.76	2.63-2.76
Multiplicity	10.0 (0.1)	9.2 (0.3)	9.5 (0.6)	9.4 (0.1)
	9.5 (0.1)	7.5 (0.2)	9.2 (1.1)	9.5 (0.2)
Completeness (%)	99.8 (0.0)	98.4 (2.3)	94.6 (8.5)	99.2 (1.1)
	100.0 (0.0)	96.6 (5.0)	90.0 (16.0)	98.7 (2.3)
R_{merge}• (%)	6.0 (1.9)	10.5 (2.4)	8.3 (2.8)	10.0 (2.0)
	46.6 (35.7)	97.1 (36.5)	77.2 (57.4)	61.1 (17.2)
$R_{p.i.m.}$• (%)	2.0 (0.6)	3.6 (0.8)	2.9 (1.0)	3.5 (0.7)
	16.0 (12.4)	37.6 (14.3)	27.7 (22.0)	20.8 (5.7)
$\langle \mathbf{I}/\boldsymbol{\sigma}(\mathbf{I}) \rangle$	26.3 (5.2)	15.6 (3.1)	21.0 (7.9)	15.9 (4.5)
	5.7 (1.9)	2.7 (0.9)	5.2 (4.4)	4.3 (2.2)
CC(1/2)•	0.9995 (0.0005)	0.9976 (0.0009)	0.9991 (0.0005)	0.9983 (0.0006)
	0.9463 (0.1072)	0.7498 (0.1497)	0.8155 (0.2309)	0.9403 (0.0408)

• See Table 3.2 footnote for definitions

Table 3.10: Student's t-test results for all data quality indicators. + represents populations that are statistically the same and - represents populations that are statistically different. Symbols in parentheses are for data in the highest resolution bin ranges defined in Table 3.7, 3.8, 3.9. C = control and H = HC-transferred.

Test Crystal	**Bacteriocin Syringacin M-1**		**Bacteriocin Syringacin M-2**		**Bacteriocin Pectocin M2**	
Treatment	C	H	C	H	C	H
Analysis	1.0 Vs 2.0	1.0 Vs 2.0	1.0 Vs 1.8	1.0 Vs 1.8	1.0 Vs 2.0	1.0 Vs 2.0
Unit Cell Parameter (a=b=) (Å)	+	+	-	-	+	+
Unit Cell Parameter (c=) (Å)	+	+	-	+	+	-
Mosaicity (°)	+	-	-	+	-	+
R_{merge}• (%)	+(+)	+(+)	-(-)	-(+)	-(-)	+(+)
$R_{p.i.m.}$ • (%)	+(+)	+(+)	-(-)	-(+)	-(-)	+(+)
I/σ(I)	+(+)	+(+)	-(-)	-(+)	-(-)	+(+)
Maximum High Resolution Limit (Å) •	+	+	-	+	-	+

• See Table 3.2 footnote for definitions

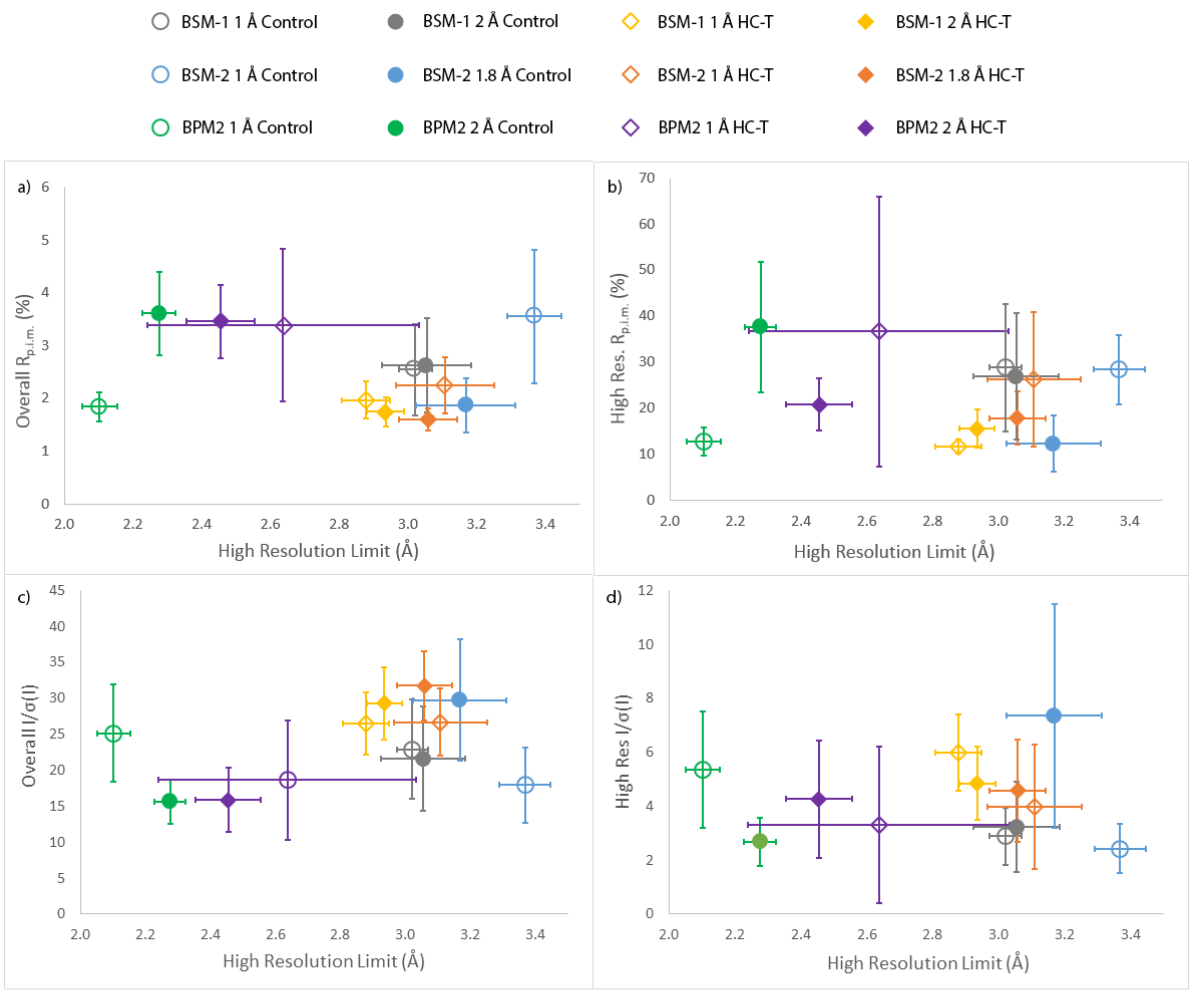


Fig. 3.18: Average overall and high resolution data qaulity indicators for like Vs like comparison of bacteriocin syringacin M-1, -2 and bacteriocin pectocin M2. Control vs Control (circles) and HC-transferred Vs HC-transferred populations (diamonds) are shwon for data collected at λ = 1 Å (hollow) and λ ≈ 2 Å (filled) plotted against average maximum high resolution limit (I/σ(I) > 2). Error bars represent one standard deviation. a) Overall $R_{p.i.m.}$. b) High resolution $R_{p.i.m.}$. c) Overall mean I/σ(I). d) High resolution mean I/σ(I).

3.4. Discussion

Previous work (see Chapter 2) demonstrated that it is possible to manipulate protein crystals in a humid stream with minimal effects on crystal integrity and data quality. However, lysozyme, thaumatin, insulin and ferritin are well characterised model samples and not representative of modern protein targets under investigation using X-ray crystallography. The samples used in Chapter 2 also diffracted close to or beyond the geometrical limits of the detector at λ ~ 1 Å, so measurement of high resolution data at longer wavelengths would be impossible. BSM, *PA*3825-EAL and BPM2 crystals diffracted to a much lower resolution and required cryo-protection prior to flash cooling, but still crystallised in a reproducible manner to allow large populations of

crystals to be examined for statistically significant differences. BSM and BPM2 crystals both have a very high solvent content, 70% and 80% respectively, and are particularly susceptible to dehydration, making them a potentially serious challenge for manipulation in the humid air stream. With *PA*3825-EAL cryoprotected in 2.4 M $ZnSO_4$ and BPM2 containing two 2Fe-2S clusters, they are both expected to have a high absorbance at longer wavelengths, making them both a good test for absorption effects with and without excess surrounding solvent.

3.4.1. Control Vs HC-transferred Samples at λ = 1.0 Å and λ ~ 2.0 Å

3.4.1.1. Bacteriocin Syringacin M

The mosaicity and unit cell parameters underwent significant changes in BSM-one and -two, with HC-transferred crystals showing an increase in size along the c-axis and a reduction in mosaicity compared to control crystals. One exception was BSM-two crystals collected at λ = 1.8 Å. These showed no significant differences in either parameter but had a higher mosaicity (0.15°) uncharacteristic of the other HC-transferred BSM samples. This suggests that BSM samples are particularly susceptible to changes induced by exposure to the humid air stream. With such a high solvent content of (~70%) it is likely that the crystal packing is being affected. This is reflected in the observed improvement in data quality indicators for HC-transferred samples in high and low resolution bins for the majority of cases taken over four different populations. Based on the pattern of reduced average mosaicity, significant or not, observed in HC-transferred samples it stands to reason that the humid air stream is having an effect on the crystals. Alternatively, the highly variable nature of the control populations may also point to an intrinsic variability in BSM crystals which makes systematic study of their data quality difficult. In either case, this makes comparing the effects of long wavelength data collection between control and HC transferred samples problematic.

3.4.1.2. *PA*3825-EAL

*PA*3825-EAL crystals show significant shrinkage in a, b and c axis unit cell parameters at λ = 1.0 Å and a and b axis unit cell parameters only at λ = 1.8 Å in control samples. The reasons behind this increase in unit cell parameters in HC-transferred samples is unknown, possibly an effect of the humid air stream or expansion of solvent or protein upon cooling/ data collection (Burmeister, 2000). However, the mosaicity of crystals and the data quality remains unaffected suggesting these changes are inconsequential.

All data quality indicators, except overall I/σ(I) show no significant difference between treatment at standard or longer wavelengths. RADDOSE-3D calculations show that the diffraction efficiency of the crystals at λ = 1.0 Å and λ = 1.8 Å is relatively unchanged, with crystals at longer wavelength diffracting more efficiently by 1.52E+11 photons/MGy. Data were collected 0.52 Å above and 0.28 Å below the zinc K edge and well below the sulphur K edge so absorption effects should be minimal at both wavelengths. Although we would expect sulphur absorption to be increased in λ = 1.8 Å, particularly controls with excess $ZnSO_4$ buffer, the effects on data quality appear to be negligible. All datasets, at both wavelengths, were subjected to the same absorption correction during scaling, namely the strict absorption correction = true and Friedel's law = false settings in XDS/XSCALE (Kabsch, 2010a, Kabsch, 2010b); with Friedel pairs treated as separate reflections when calculating absorption correction factors.

3.4.1.3. Bacteriocin Pectocin M2

BPM2 crystals reacted very poorly to manipulation in the humid air stream, showing a consistent shrinkage in the a and b unit cell parameters and an increase in mosaicity. This has a clear knock on effect, significantly reducing the high resolution limit and negatively affecting the quality of diffraction data from the crystals. As a result, comparison of control and HC transferred samples at

different wavelengths yielded no useful information on the effects of data quality when collecting at long wavelengths with or without excess solvent.

3.4.2. Like Vs Like Treatments at λ = 1.0 Å and λ ~ 2.0 Å

3.4.2.1. Bacteriocin Syringacin M

Comparison of data from like treatments (i.e. control vs control, etc.) collected at different wavelengths is somewhat more reliable in BSM-one compared to BSM-two. BSM-one control data at λ = 1.0 Å or λ = 2.0 Å show no significant difference in unit cell parameters, mosaicity or other data quality indicators except mosaicity in the HC-transferred crystals, which is increased in the λ = 2.0 Å populations. BSM-two data on the other hand are much more variable, with control crystals demonstrating a shrinkage in all unit cell parameters and an increase in mosaicity in the λ = 1.0 Å data. BSM-two HC-transferred crystals also see a shrinkage in the a and b axis of λ = 1.0 Å crystals.

The stability of BSM-one's unit cell and mosaicity is reflected in data quality indicators which show no differences when comparing like with like treatments at different wavelengths. The same cannot be said for BSM-two data, with control crystals showing a significant improvement in data quality at λ = 1.8 Å. The similarity in BSM-one populations and the improvements in BSM-two data when going to longer wavelengths were unexpected based on predictions by RADDOSE, which suggests a 8.5 fold drop in diffraction efficiency with increasing wavelengths from λ = 1.0 Å to λ = 2.0 Å (Fig. 3.9b). Based on the wide range of differences observed in crystal integrity measures in control and HC-treated crystals BSM appears to be an inappropriate crystal for used in these systematic studies based on the intrinsic variability of the crystals themselves. This variability emphasises the need for robust statistical analysis in method development for X-ray crystallography. Experimentation on a single BSM crystal from control and HC-transferred crystals, from either BSM-1 or BSM-2, could have resulted in any number of conclusions on the effects of excess surrounding

solvent on data quality at long wavelengths depending on the particular crystal data were collected from.

3.4.2.2. Bacteriocin Pectocin M2

Although control samples of BMP2 show a significant improvement over HC-transferred samples, the changes caused by the humid air stream are reproducible; allowing comparison of like with like treatments, i.e. HC-transferred vs HC-transferred, at λ = 1.0 Å and λ = 2.0 Å. Control populations processed to the same high resolution limit show a significant improvement in data quality indicators when collected at λ = 1.0 Å. No such differences in data quality indicators are observed between HC-transferred populations at λ = 1.0 Å and λ = 2.0 Å; suggesting these effects stem from the presence of excess surrounding solvent. The HC-transferred populations collected at λ = 1.0 Å show a large standard deviation which could skew calculations of p-values, however, the three datasets causing this have not been identified as outliers by the Grubb's test and removing the datasets has no effect on the final result of the t-test.

BPM2 contains a 2Fe-2S cluster (Grinter et al., 2014) which will greatly increase the absorption of X-rays at λ = 2.0 Å, however, these effects appear insignificant in populations where excess surrounding solvent has been removed. RADDOSE-3D calculations suggest the diffraction efficiency of crystals in the optimum orientation should dramatically decrease from 5.78E+12 photons/MGy at λ = 1.0 Å to 6.25E+11 photons/MGy at λ = 2.0 Å, a 8.2 times increase in efficiency. However, RADDOSE-3D does not take excess surrounding solvent into account so we would expect the same decrease in photons/MGy in both control and HC-transferred samples. Calculations with buffer alone show the cryo buffer would receive a smaller maximum dose than the crystal in 180° collections suggesting the absorption by excess surrounding buffer is not sufficient to cause the significant differences observed in control data quality at λ = 2.0 Å.

These differences in data quality may also be the result of radiation damage experienced through increased absorption of X-rays at λ = 2.0 Å. Primary radiation damage occurs by direct interaction of the X-ray beam with electrons to generate photoelectrons and ejection of energetic electrons from atoms (Teng and Moffat, 2000), with the first port of call for site specific radiation damage being reduction of metallic centres (Garman and Weik, 2015, Hough et al., 2008a, Ellis et al., 2008, Yano et al., 2005b). This makes BMP2's 2Fe-2S centre a likely target as an electron sink for the free electrons (Ellis et al., 2002, Nishino and Okamoto, 2000). However, this damage should occur regardless of the presence of excess surrounding solvent once the appropriate dose limit has been reached (Owen et al., 2006a, Murray et al., 2005, Henderson, 1990). A change in colour of BPM2 crystals is clearly seen in both control and HC-transferred samples (Fig. 3.19), signifying the 2Fe-2S cluster has been reduced. Despite this, no reduction in data quality is observed in HC-transferred populations while control populations collected at λ = 2.0 Å do show a significant decrease in data quality, suggesting that primary radiation damage is not the cause.

Secondary damage results from the generation of free radicals which will vary by the nature of the solvent (Murray et al., 2004, Shimizu et al., 2007), temperature (Burmeister, 2000, Owen et al., 2006a, Ravelli and McSweeney, 2000) and presence of scavengers (Allan et al., 2013, Murray and Garman, 2002). Radiation damamge presents as either global or specific damage, with global radiation damage causing a loss of high resolution reflections and an increase in unit cell volume, $R_{merge/meas/p.i.m}$, B-factors and mosaicity (Garman, 2010). Specific damage occurs between covalent bonds in a specific reproducible order in cryo-crystallography; with disulphide bridges breaking first, followed by decarboxyation of glutamates and aspartates, loss of hydroxyl groups from tyrosines and finally the cleavage of carbon sulphur bonds in methionines (Garman, 2010).

All populations were collected at the same temperature and no free radical scavengers were present in the buffers so solvent presents the only variable left to investigate. The process of secondary damage requires thermal diffusion of these reactive atomic and molecular radicals, which

is greatly reduced at cryo temperatures (Ravelli and McSweeney, 2000, Garman, 2003b). However, mobile hydroxyl radicals and electrons have been reported at 77K (Lange et al., 1995, Whalley, 1973, Eiben and Taub, 1967), along with other accounts of secondary radiation damage at cryogenic temperatures (Sevilla and D'Arcy, 1978, Sevilla et al., 1979, Box and Budzinski, 1971). BPM2 has a high solvent content of ~ 80% and the presence of excess surrounding solvent will greatly increase the number of solvent and water molecules available to form hydroxyl radicals, hydrated electrons, protons, hydrogen radicals and hydrogen gas (Faraggi et al., 1995, Alexander and Lett, 1967, Weik et al., 2000, Burmeister, 2000, von Sonntag and Schuchmann, 1994). The radiolysis of water molecules is depicted in figure 3.20. Ionisation produces a free electron and a water cation leading to production of hydroxyl radicals (Hiroki et al., 2002). The damage caused by these highly reactive free radical species may account for the detrimental effects on data quality observed in λ = 2.0 Å control populations.

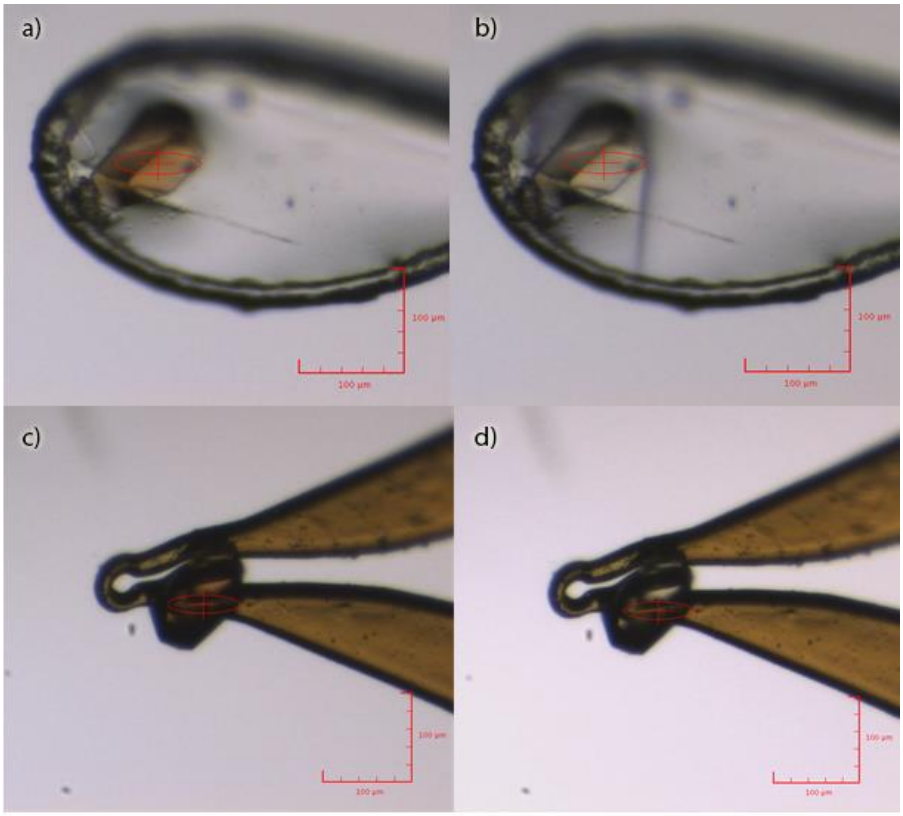


Fig. 3.19: Bacteriocin pectocin M2 control (a-b) and HC-transferred (c-d) crystals before (left) and after (right) complete (180° rotation) data collections at λ = 1.0 Å. Note the 'bleaching' of the crystals resulting from the exposure to the X-ray beam. The same average X-ray dose (0.22 MGy, as calculated by RADDOSE-3D (Zeldin et al., 2013)) was applied in each case.

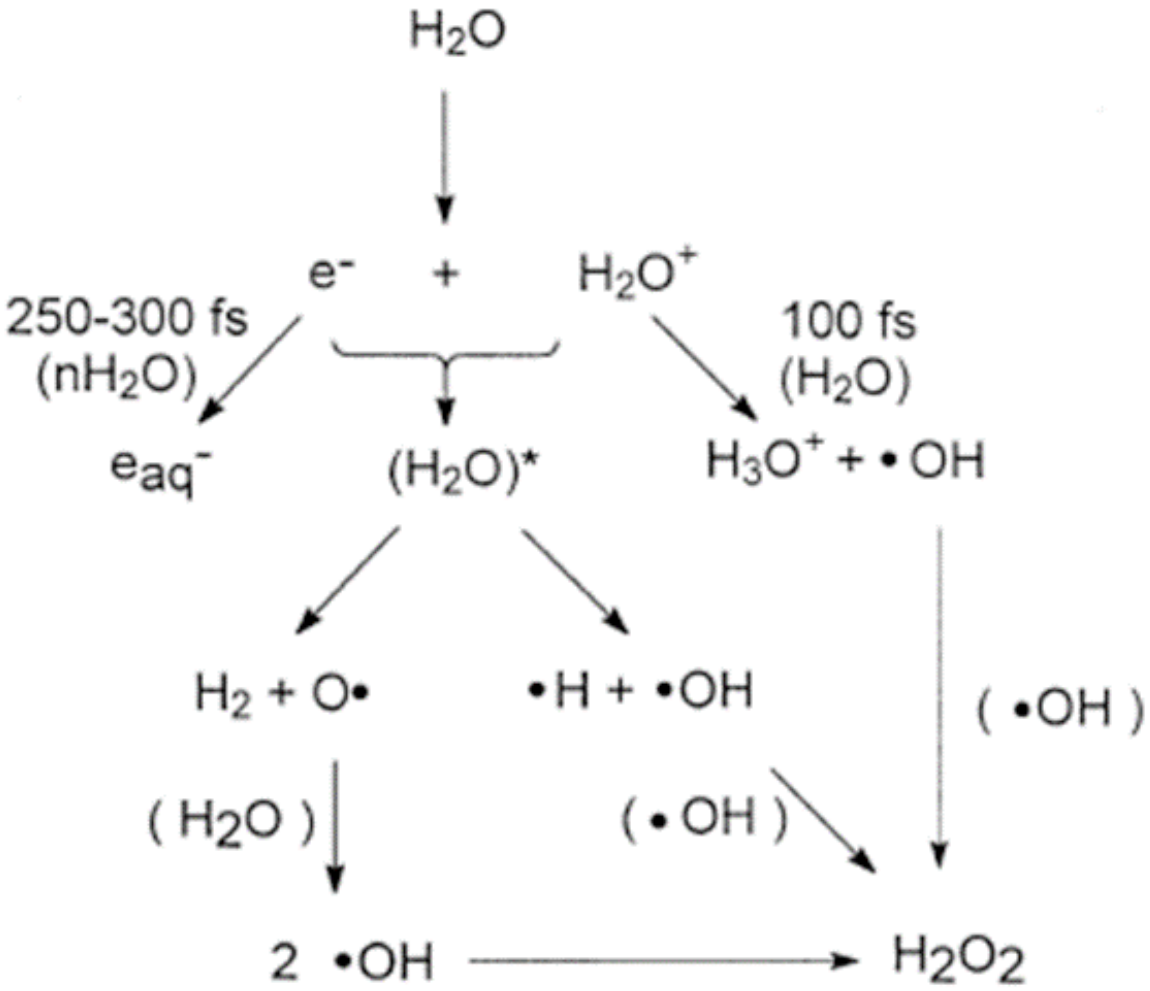


Fig. 3.20: Radiolysis of water at room temperature and neutral pH to form hydrogen peroxide. Water is ionised to produce an electron (e^-) and a molecular cation of water (H_2O^+), followed by either electron hydrtaion (e_{aq}^-, left) and proton transfer by the cation (right) or dissociative recombination (centre). Both paths produce •OH radicals for formation of hydrogen peroxide. Figure taken from Hiroki et al. (2002).

3.5. Conclusions

This work has built on the previous study presented in Chapter 2, in which a humidity controlled device and micromanipulators were employed to allow precise removal of surrounding solvent from protein crystals, in preparation for either transfer to a different mount, for in-vacuum experiments, or for crystallographic study. Following those findings - that robust 'model' protein crystals can be manipulated routinely without causing deterioration of their diffraction quality - the

next step was to conduct experiments on more “true to life” protein crystals (weaker diffracting, limited diffraction limit, high solvent content, physically more fragile, etc) and statistically assess the effects of crystal manipulation and excess solvent on their data quality, including at longer wavelengths.

As expected, protein samples with higher solvent content showed some variability when exposed to the humid air stream. BSM crystals proved highly variable, with four populations of control and HC-transferred crystals tested, giving different results on the state of mosaicity and unit cell parameters both between treatments and when comparing like treatments. The knock on effect of these changes in crystal integrity on data quality were evident for HC-transferred crystals which, in the majority of cases, showed an improvement in high resolution limit and other data quality indicators. While an improvement in data quality in the humid air stream can be considered a favourable result, both the HC-transferred and control crystals demonstrated too much variability to give a clear impression of the effects of longer wavelengths and excess solvent on data quality. Changes to the BSM crystals in the humid air stream suggests that the optimum humidity may not have been determined, despite repeated experiments producing a relative humidity within 0.5%. It may be that a more accurate means of determining the relative humidity of BSM crystals than the ‘loop method’ is required to optimise conditions for BSM to become a more reliable sample for systematic studies.

*PA*3825-EAL protein crystals proved much more stable in the humid air stream. Although significant differences were observed in some unit cell axis upon exposure to the humid air stream, up to 0.3 Å the mosaicity remained unaffected. With a solvent content of approximately 50%, *PA*3825-EAL is expected to be less susceptible to the changes seen in the more highly hydrated BSM crystals. The changes in unit cell parameters between treatments appears to have had no knock on effects on the quality of diffraction data obtained. The presence of excess solvent also appears to have no significant effects on data quality at λ = 1.8 Å. This may be explained by the predicted

diffraction efficiency from RADDOSE which suggests little difference between crystals collected at λ = 1.0 Å and λ = 1.8 Å. RADDOSE predicts a peak in diffraction efficiency for *PA*3825-EAL around λ = 1.4 Å, however, this is likely the result of the zinc absorption edge due to the 2.4 M zinc sulphate present as a cryo-protectant.

BPM2 presents an interesting case of crystals undergoing a negative change in data quality following exposure to the humid air stream in a highly reproducible manner that allows further study. At its best, BPM2 crystals diffracted to an average maximum resolution of 2.16 Å (λ = 1.0 Å control crystals); following exposure to the humid air stream this dropped considerably to an average of 2.64 Å (λ = 1.0 Å HC-transferred crystals). Despite this drop in high resolution limit, data quality in HC-transferred samples remains high with reasonable merging statistics. The more interesting conclusions appear when comparing populations handled in the same manner (i.e. control vs control) at different wavelengths. For HC-transferred naked protein crystals at λ = 1.0 Å and λ = 2.0 Å there is no observable difference in data quality indicators from a total population of 26 protein crystals. Based on the 9.2-fold drop in diffraction efficiency from data collection at λ = 1.0 Å to λ = 2.0 Å, as calculated by RADDOSE, this should be reflected in average data quality indicators. Although this is not observed in HC-transferred populations it is seen in control populations, where λ = 1.0 Å samples show significantly superior data quality indicators than at λ = 2.0 Å. If this decrease in data quality was caused solely by using long wavelength X-rays we would expect HC-transferred samples to show a similar decrease in data quality. As this is not observed, the only other variable is the presence of excess solvent which, when paired with long wavelength X-rays, leads to an increase in absorption, background scatter and the production of high energy radicals which have detrimental effects on data quality.

Although BSM proved incompatible with the systematic study of data quality in relation to data quality it does provide insight into the use of the HC1 in sample handling. Treatment with the HC1 is highly dependent on the behaviour of protein crystals when removed from the stable

environment of the mother liquor, namely the degree of variability in crystal packing. Of all the protein crystals tested in Chapters 2 and 3, BSM has the highest solvent content at ~ 80%. Initial crystals studied (Chapter 2) had solvent contents between 26% and 60%, with the two higher of the four, thaumatin (57%) and ferritin (60%), demonstrating statistically significant variability in unit cell parameters and mosaicity. However, the data quality indicators remained unaffected (Table 2.7). This chapter looked to investigate more "true to life" protein crystals, with BSM, *PA*3825-EAL and BPM2 having solvent contents of 80%, 50% and 70%, respectively. *PA*3825-EAL interacted with the humid air stream in a similar manner to thaumatin and ferritin, with statistically significant changes in unit cell parameters while data quality remained unaffected (Table 3.6). BSM and BPM on the other hand demonstrated significant changes in unit cell parameters and mosaicity which had a knock on effect on data quality between control and HC-transferred samples as discussed above in detail. Based on the observations of crystal variability, both crystal integrity and data quality, and its relation to solvent content a scale of difficulty guideline for use of the HC1 for manipulation of protein crystals can be proposed; with solvent contents up to 50% as relatively easy cases, 51-60% as potentially problematic cases and >61% as complicated/difficult cases.

Based on this study, which has used statistical analysis based on multiple crystal experiments (and crystal types), it is not possible to produce a more specific predictive model for the effects on data quality of excess solvent paired with long wavelength X-rays. A number of variables including the buffers used for cryo-protection, crystal size and orientation, solvent content and the presence of metal centres and other heavy atoms could have a significant effect on the final data from a diffraction experiment. Nevertheless, crystal manipulation and optimising conditions for data collection will become a more important issues for expermintalists as we approach much longer wavelengths – for example, 3-4 Å proposed for I23 at Diamond - with a view to perform S-SAD phasing experiments. For this study, investigations using these long wavelengths was not yet technically possible and must remain for the future.

3.6. Appendix

3.6.1. Dose Calculations

Table 3.A1: Calculated fluxes at different energies from Diamond Light Source beamline I04 from data measured at 100% transmission at 300 mA ring current (Fig. 3.5).

Energy (keV)	Wavelength (Å)	Flux Ph/s)
12.4	1.0	2.00E+10
10.3	1.2	3.16E+10
8.9	1.4	3.16E+10
7.8	1.6	2.51E+10
6.9	1.8	2.00E+10
6.2	2	1.26E+10
5.6	2.2	1.26E+10
5.2	2.4	1.00E10

Table 3.A2: RADDOSE-3D input parameters for bacteriocin syringacin M, *PA*3825-EAL and bacteriocin pectocin M2 at different X-ray energies. Energies and fluxes are listed in Table A3.1. Numbers in brackets represent simulations for bacteriocin pectocin M2 crystal orientation two (Fig. 3.6b).

Protein	**Bacteriocin syringacin M**		***PA*3825-EAL**		**Bacteriocin Pectocin M2**	
Crystal/Buffer	Crystal	Buffer	Crystal	Buffer	Crystal 1 (2)	Buffer
Beam Parameters						
Full Width Half Max. (µm)	90 x 45	90 x 45	90 x 45	90 x 45	90 x 45	90 x 45
Crystal Parameters						
Xtal Dimensions (µm)	185 x 160 x 140	185 x 160 x 140	100 x 100 x 100	100 x 100 x 100	150 x 250 x 150 (250 x 150 x 150)	150 x 250 x 150 (250 x 150 x 150)
Number of Monomers	1	0	1	0	2	0
Number of Residues	284	284	259	259	279	279
Solvent Heavy Atoms	Ca 30, Cl 30, Mg 30, Cl 30	Ca 30, Cl 30, Mg 30, Cl 30	Zn 2400, S 2400, Na 100	Zn 2400, S 2400, Na 100	S 2000, S 100	S 2000, S 100
Protein Heavy atoms	S 7, Ca 1	N/A	S 7	N/A	Fe 4, S 9	N/A
Solvent Fraction (%)	79.51	100	48.76	0	70.42	0
Wedge Parameters						
Angular Range (°)	0-120	0-120	0-120	0-120	0-180	0-180
Exposure Time (s)	120	120	120	120	180	180

3.6.2. Test for Normality

Table 3.A3: Correlation coefficients to test for normality in control and HC-transferred populations of bacteriocin syringacin M, *PA*3825-EAL and bacteriocin pectocin M2 crystals collected at λ = 1Å. Critical values are presented for α = 0.01, (α = 0.05) and [α = 0.10]. Populations are considered highly normal if correlation coefficients exceed all critical values, normal if higher than α = 0.01 and α = (0.05) critical values and weakly normal if only higher than α = 0.01 critical values.

Protein	**Bacteriocin syringacin M-1**		**Bacteriocin syringacin M-2**		***PA*3825-EAL**		**Bacteriocin Pectocin M2**	
Control/ HC-trans.	Control	HC-trans.	Control	HC-trans.	Control	HC-trans.	Control	HC-trans.
Overall R_{merge} (%)	0.92	0.96	0.95	0.94	0.92	0.95	0.97	0.98
High Res. R_{merge} (%)	0.95	0.97	0.84	0.89	0.93	0.97	0.98	0.95
Overall $R_{p.i.m.}$ (%)	0.91	0.96	0.95	0.94	0.93	0.95	0.87	0.97
High Res. $R_{p.i.m.}$ (%)	0.95	0.97	0.84	0.88	0.93	0.94	0.97	0.95
Overall I/σ(I)	0.97	0.95	0.96	0.95	0.94	0.96	0.93	0.96
High Res. I/σ(I)	0.98	0.96	0.96	0.96	0.93	0.99	0.95	0.94
High Res. Limit (Å)	0.98	0.97	0.91	0.96	0.95	0.90	0.98	0.97
Unit Cell Parameters (a=b=)	0.99	0.97	0.98	0.91	0.97	0.98	0.91	0.90
Unit Cell Parameters (c=)	0.98	0.95	0.96	0.82	0.97	0.97	0.82	0.92
Mosaicity (°)	0.96	0.99	0.97	0.95	0.99	0.96	0.97	0.89
Critical Value	**0.86 (0.91) [0.92]**	**0.86 (0.91) [0.92]**	**0.85 (0.90) [0.92]**	**0.84 (0.89) [0.91]**	**0.89 (0.92) [0.94]**	**0.89 (0.92) [0.94]**	**0.90 (0.94) [0.95]**	**0.91 (0.94) [0.95]**

Table 3.A4: Correlation coefficients to test for normality in control and HC-transferred populations of bacteriocin syringacin M, *PA*3825-EAL and bacteriocin pectocin M2 crystals collected at λ ~ 2Å. Critical values are presented for α = 0.01, (α = 0.05) and [α = 0.10]. Populations are considered highly normal if correlation coefficients exceed all critical values, normal if higher than α = 0.01 and α = (0.05) critical values and weakly normal if only higher than α = 0.01 critical values.

Protein	**Bacteriocin syringacin M-1**		**Bacteriocin syringacin M-2**		***PA*3825-EAL**		**Bacteriocin Pectocin M2**	
Control/ HC-trans.	Control	HC-trans.	Control	HC-trans.	Control	HC-trans.	Control	HC-trans.
Overall R_{merge} (%)	0.94	0.96	0.97	0.97	0.94	0.96	0.97	0.95
High Res. R_{merge} (%)	0.96	0.96	0.96	0.97	0.98	0.97	0.98	0.97
Overall $R_{p.i.m.}$ (%)	0.95	0.91	0.96	0.97	0.96	0.97	0.96	0.95
High Res. $R_{p.i.m.}$ (%)	0.96	0.97	0.96	0.97	0.94	0.96	0.97	0.97
Overall I/σ(I)	0.98	0.98	0.96	0.98	0.97	0.98	0.99	0.95
High Res. I/σ(I)	0.98	0.93	0.98	0.95	0.97	0.87	0.98	0.98
High Res. Limit (Å)	0.95	0.98	0.99	0.96	0.97	0.93	0.98	0.98
Unit Cell Parameters (a=b=)	0.99	0.95	0.95	0.95	0.98	0.98	0.99	0.87
Unit Cell Parameters (c=)	0.97	0.97	0.96	0.93	0.94	0.96	0.95	0.92
Mosaicity (°)	0.97	0.97	0.93	0.94	0.93	0.95	0.87	0.96
Critical Value	**0.88 (0.92) [0.93]**	**0.88 (0.92) [0.93]**	**0.86 (0.91) [0.92]**	**0.86 (0.91) [0.92]**	**0.89 (0.92) [0.94]**	**0.89 (0.92) [0.94]**	**0.89 (0.93) [0.94]**	**0.89 (0.92) [0.94]**

Chapter 4 – Cancerous Inhibitor of Protein Phosphatase 2A, a Target for S-SAD phasing

As discussed in previous chapters (1, 2 and 3), X-ray crystallography has developed considerably from its humble beginnings solving the structure of NaCl (Bragg and Bragg, 1913), and is now tackling more and more difficult targets. Thanks to improvements in data quality through more stable high brilliance synchrotron sources, improved phasing software and developments in detector technology (Rajendran et al., 2011b) there has been a resurgence in the use of S-SAD to solve the phase problem (Chapter 1). Cancerous Inhibitor of Protein phosphatase 2A (CIP2A) is a recently discovered oncoprotein prevalent in a wide number of cancers. Currently no structural data for CIP2A is available beyond sequence based secondary structure predictions. In the event of purification, crystallisation and data collection CIP2A has no homology models for phasing by molecular replacement, however, full length CIP2A has 37 sulphur atoms (21 Cys and 16 Met) making it a potential target for S-SAD phasing. Full length CIP2A, and the N-terminal and C-terminal recombinant CIP2A proteins in this study have Bijvoet ratios of 0.33%, 0.38% and 0.22% at λ = 1.0 Å, all very difficult cases for S-SAD phasing. By increasing the wavelength to 2 Å, 3 Å or even 4 Å the bijvoet ratio increases by approximately 4.1, 8.0 and 12.8 fold, respectively, making successful phasing much more likely. This project aimed to express, purify and crystallise these recombinant protein fragments of CIP2A N- and C-terminal regions for use on the long wavelength X-ray crystallography beamline (I23) currently under development at Diamond Light Source.

4.1. Cancer Biology

Understanding and treating cancer remains one of the biggest challenges facing modern medicine. Tumourigenesis is a complex process requiring cells to be able to sustain proliferative signalling, evade growth suppression, resist cell death, enable replicative immortality, inducing angiogenesis and activate invasion and metastasis. These are usually referred to as the “hallmarks of cancer” (Fig. 4.1) (Hanahan and Weinberg, 2000). Since 2000 these hallmarks have been expanded to include the emerging ones of deregulating cellular energetics and avoiding immune destruction, as well as the cancer cell enabling characteristics of genome instability and mutation and tumour-

promoting inflammation (Fig. 4.1) (Hanahan and Weinberg, 2011). The multistep process cells go through to acquire these hallmark capabilities, which may occur in any order each by a variety of mechanisms, illustrates the complexity that clinicians and health sector scientists are faced with when treating and researching new treatments for cancers. Examples of potential drugs to target individual hallmarks are shown in Figure 4.1, some of which have entered clinical trials with a select few approved for clinical use.

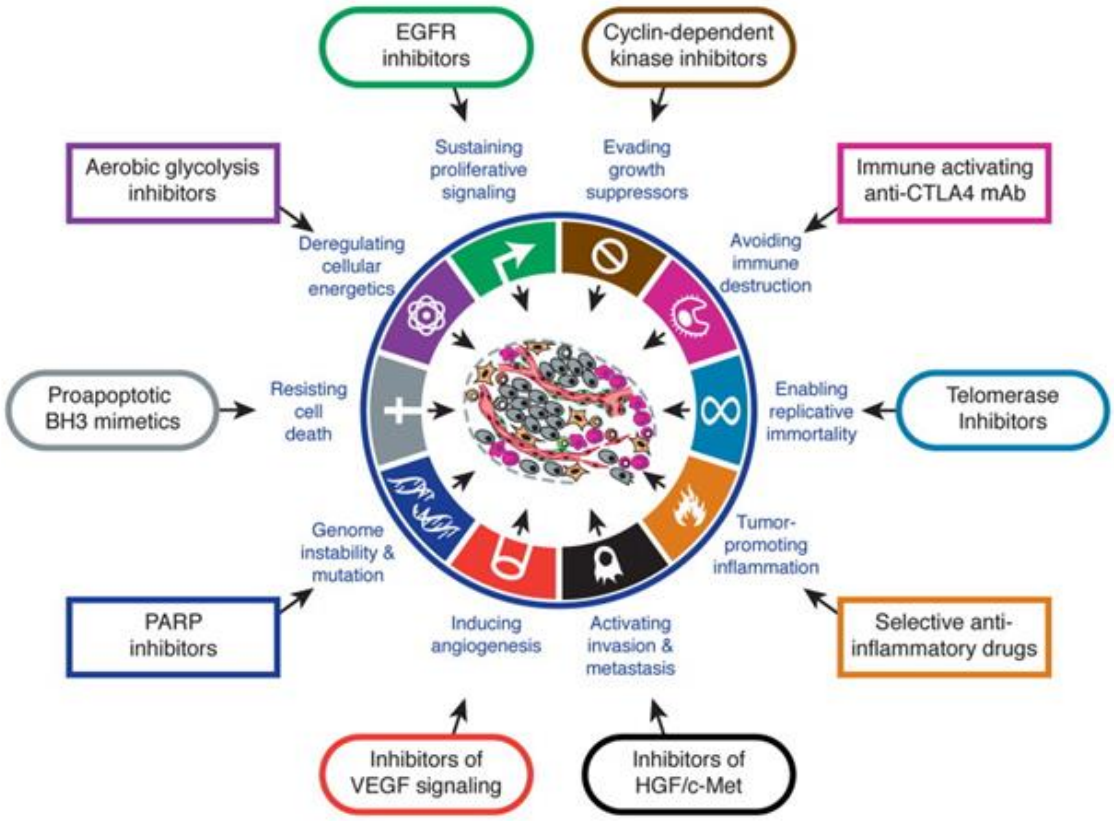


Fig. 4.1: The hallmarks of cancer, including the emerging hallmarks (deregulating cellular energetics and avoiding immune destruction) and enabling characteristics (genome instability/ mutation and tumor-promoting inflammation. The outer ring represents potential drug targets for each aquired ability in cancer cells. Figure taken from Hanahan and Weinberg 2011.

Phosphoregulation of proteins is one of the most common mechanisms for regulating protein function and cell signalling, including cell growth, survival and differentiation (Westermarck and Hahn, 2008, Khanna et al., 2013, Ciccone et al., 2015). As a result, phosphorylation and

dephosphorylation are subject to tight controls by complementary switches in the form of kinases and phosphatases. Deregulation of this tightly controlled balance by upregulating protein kinases or down regulating protein phosphatases has been attributed to pathogenesis of a wide variety of human diseases, including various cancers (Westermarck and Hahn, 2008, Eifert and Powers, 2012, Perrotti and Neviani, 2013). Interestingly, it has been found that up regulation of kinase signalling via the RAS pathway cannot cause transformation without simultaneous inhibition of protein phosphatase 2A (PP2A) (Westermarck and Hahn, 2008, Perrotti et al., 2010). This points to PP2A reactivation as a potential therapy in conjunction with kinase inhibition (Westermarck and Hahn, 2008, Perrotti and Neviani, 2013). A key factor in inhibition of PP2A is the overexpression of a recently discovered oncoprotein, cancerous inhibitor of protein phosphatase 2A (CIP2A).

4.1.1. Cancerous Inhibitor of Protein Phosphatase 2A

The protein "p90" was first discovered through association of "p90" auto-antibodies with p62, a cytoplasmic RNA binding protein involved in regulation of IGF II mRNA expression (Soo Hoo et al., 2002), and later named CIP2A when identified as an endogenous PP2A-interacting protein (Junttila et al., 2007). Junttila et al also identified CIP2A as oncogenic PP2A inhibitor by its ability to replace small-t antigen in the transformation of p53 and Rb tumour suppressor deficient HEK cells. Since its discovery, CIP2A has been linked to a plethora of cancer processes including decreased cell viability, anchorage independent growth, progenitor cell self-renewal, protection from senescence induction, cell cycle regulation and mitosis (Khanna et al., 2013). More importantly, depletion of CIP2A with small interfering RNAs (siRNAs) in xenografted tumours has been shown to inhibit growth in various cancers (Junttila et al., 2007, Come et al., 2009, Xue et al., 2013) and hypomorphic CIP2A mice models are healthy and viable (Laine et al., 2013); although spermatogenesis is affected (Ventela et al., 2012).

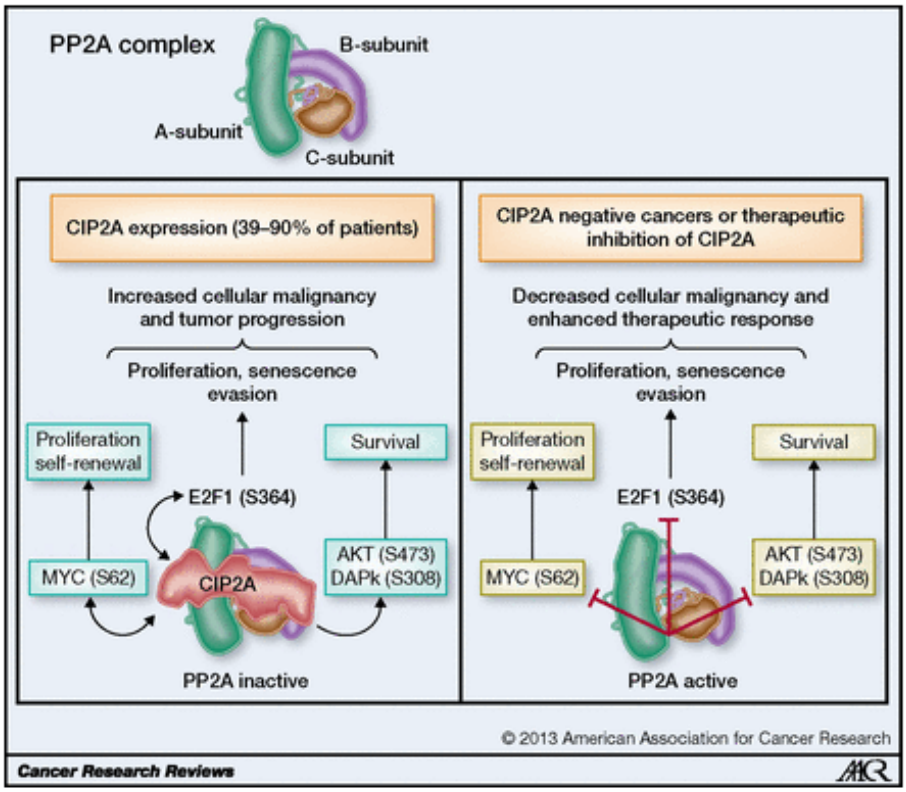


Fig. 4.2: Cartoon representation of the oncogenic mechanisms of CIP2A expression. MYC, E2F1, AKT and DAPk remain phosphorylated at labelled sites in the presence of CIP2A over expression, leading to tumorigenesis (left). Inhibition of CIP2A expression and/or activity prevents proliferation, senescence and evasion of treatment and the immune system (right) by allowing dephosphorylation of target proteins by PP2A. Double sided arrows a indicate positive feedback loop. Figure adapted from Khanna et al. (2013).

CIP2A expression has been observed in a wide variety of cancers, both solid and haematological, including gastric, bladder, ovarian, tongue, hepatocellular, colon, non-small cell lung carcinoma and chronic myeloid leukaemia; with CIP2A overexpression found in 39-90% of patient samples (Khanna et al., 2013). In addition to being highly prevalent, CIP2A expression has been correlated with tumour grade (Come et al., 2009, Dong et al., 2011, Khanna et al., 2009, Xue et al., 2013, Laine et al., 2013) and negative response to treatment (Laine et al., 2013, Bockelman et al., 2011). CIP2A acts through several common signalling pathways including Myc, the EGFR-MEK-ETS1 pathways, p53 inactivation and over expression of E2F1 and ATF2 (Fig. 4.4) (Khanna et al., 2013).

4.1.1.1. Myc, CIP2A and Cancer

Myc is a transcription factor for the leucine zipper family that has been linked to cell cycle regulation, proliferation, growth, differentiation and metabolism; unsurprisingly, deregulation of Myc has been found to promote cell transformation and tumour progression (Junttila and Westermarck, 2008). Phosphorylation plays a key role in the regulation of Myc, with phosphorylation of Serine 62 (S62) by ERK stabilising Myc, and phosphorylation of Threonine 58 (T58) by GSK-3β targeting Myc for degradation (Pulverer et al., 1994, Sears et al., 2000). However, degradation of Myc by ubiquitination is dependent on dephosphorylation of S62 by PP2A (Yeh et al., 2004). *Ergo*, expression of CIP2A prevents PP2A from dephosphorylating Myc S62 leading to cell proliferation and transformation (Fig. 4.3) (Junttila and Westermarck, 2008). In addition to preventing Myc degradation, CIP2A has been found to promote expression of Myc which, in turn, stimulates CIP2A mRNA and protein expression to form a positive feedback loop (Kerosuo et al., 2010, Mathiasen et al., 2012).

4.1.1.2. E2F1, CIP2A and Cancer

E2Fs are a family of transcription factors that can act as transcriptional activators (E2F1-3a) and transcriptional repressors (E2F3b/E2F4-8) (Putzer and Engelmann, 2013). E2F1 has been found to induce apoptosis, acting as an antitumourigenic fail-safe in a similar way to p53 (Putzer and Engelmann, 2013); although E2F1 can act by p53 dependent and independent mechanisms (Stanelle and Pützer, 2006). As a result E2F1 was championed as a predictive factor of radiosensitivity and chemosensitivity (Lee et al., 2008). However, it has since been found that increased E2F1 along with reduced retinoblastoma protein 1 can be associated with resistance to chemotherapy (Han et al., 2003, Sharma et al., 2010). E2F1 and CIP2A act as part of an oncogenic positive feedback loop with E2F1 activating the CIP2A promoter and CIP2A preventing removal of a phosphate from Serine 364 of E2F1 which may explain this (Laine et al., 2013). This leads to reduced activity of p53 and inhibition of senescence induction.

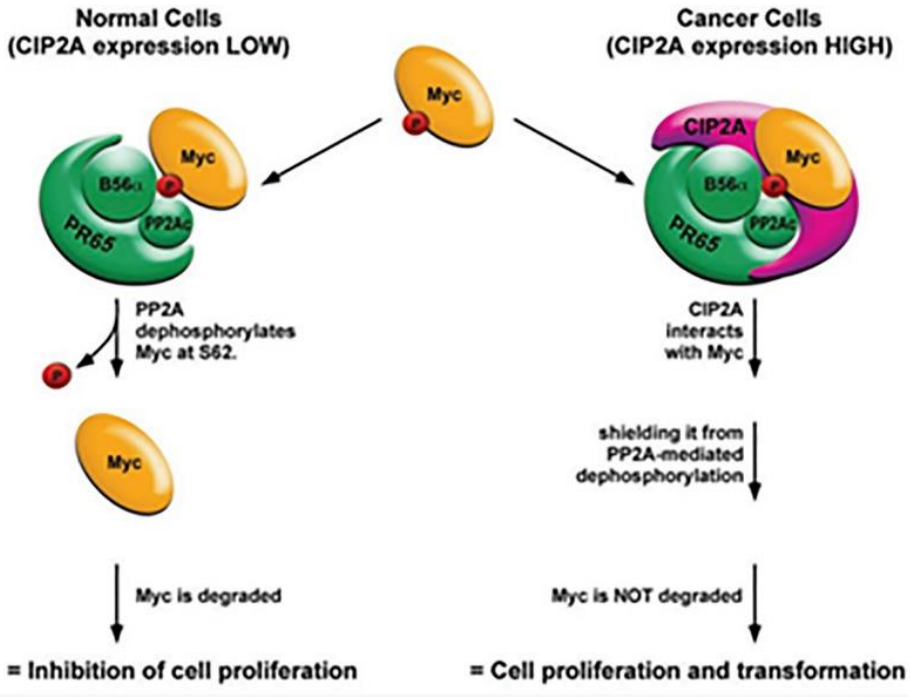


Fig. 4.3: Myc stabilisation as mediated by direct binding of CIP2A to Myc. Normal cells show degradation of Myc by ubiquitination (left) while excess Myc, due to CIP2A preventing PP2A-mediated dephosphorylation, causes cell proliferation and transformation (Right). Figure adapted from Junttila and Westermarck 2008.

4.1.1.3. Akt, CIP2A and Cancer

Akt is a Serine/Threonine kinase from the AGC family with 3 isoforms (Akt1/PKBα, Akt2/PKBβ and Akt3/PKBγ). Akt recognises the motif R-X-R-X-X-S/T-B (Alessi et al., 1996), where X represents any amino acid and B is a bulky hydrophobic residue, to allow direct phosphorylation of various cellular targets involved in preventing apoptosis, promoting protein expression and regulating cellular metabolism (Wang et al., 2015). Cellular targets for activation and inhibition by Akt and the affected cellular processes are shown in Figure 4.4. Akt itself is activated through the PI3K pathway and is further regulated through dephophoprylation of threonine 308 and serine 473 by PP2A to switch between being anti- and pro-apoptotic (Andrabi et al., 2007). Overexpression of CIP2A prevents dephosphorylation of these sites to promote cell survival.

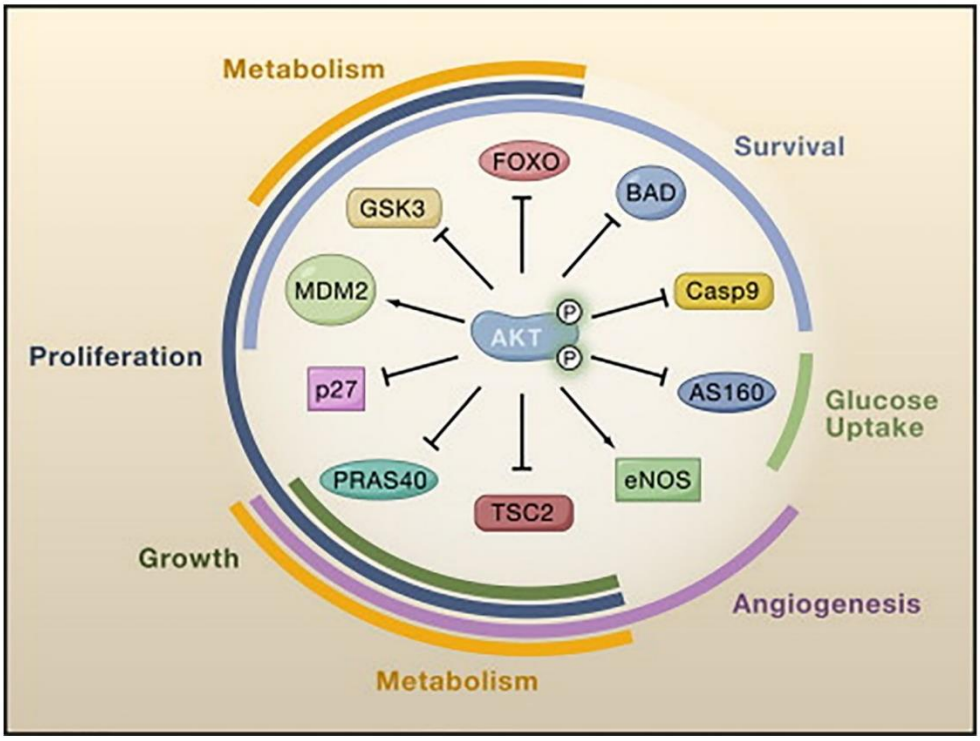


Fig. 4.4: A schematic of PI3K activated Akt and the resulting Akt-mediated phosphorylation for activation and inhibition of various proteins, as specified, and the affected cellular processes from Akt-mediated regulation. Figure adapted from Manning and Cantley (2007).

4.1.2. CIP2A as a Drug Target

Based on the evidence that transformation of cells requires the activation of kinases and suppression of PP2A, and the fact that PP2A itself is rarely mutated; targeting PP2A inhibitors presents a feasible option for cancer therapy (Khanna et al., 2013). The widespread influence of CIP2A on multiple pathways, as discussed above, and the evidence that mouse models with depleted CIP2A are healthy suggests CIP2A makes a good target for treating many cancers. How drugable CIP2A is remains unclear due to the lack of any structural data or evidence for enzymatic activity. Although the structure of CIP2A remains elusive, bioinformatics analysis has provided some insight into the possible structure and domain organisation.

4.1.3. Protein Phosphatase 2A

The prevalence of CIP2A in cancer is thought to be the result of its interaction with PP2A (Junttila et al., 2007), an important regulator in a host of cellular processes. PP2A represents a family of heterotrimeric serine/threonine phosphatases responsible for a large proportion of serine/threonine phosphatase activity in eukaryotes (Millward et al., 1999). PP2A consists of a 36 KDa catalytic subunit ($PP2A_C$, Fig. 4.5 purple) and a 65 KDa structural subunit ($PP2A_A$, Fig. 4.5 green), each of which have α and β isoforms (Aα, Aβ, Cα and Cβ), as well one of a variety of regulatory B subunits ranging from 50-135 KDa (Fig. 4.5 cyan) (Cohen, 1990, Cho and Xu, 2007). Figure 4.5 is pictured with the regulatory subunit B56γ1, a proposed tumour suppressor consisting of eight pseudo-HEAT repeats (Cho and Xu, 2007). The combination of different A, B and C subunit isoforms gives a total of 75 possible distinct PP2A holoenzymes which can affect activity as well as determine substrate and tissue specificity (Cegielska et al., 1994, Eichhorn et al., 2009).

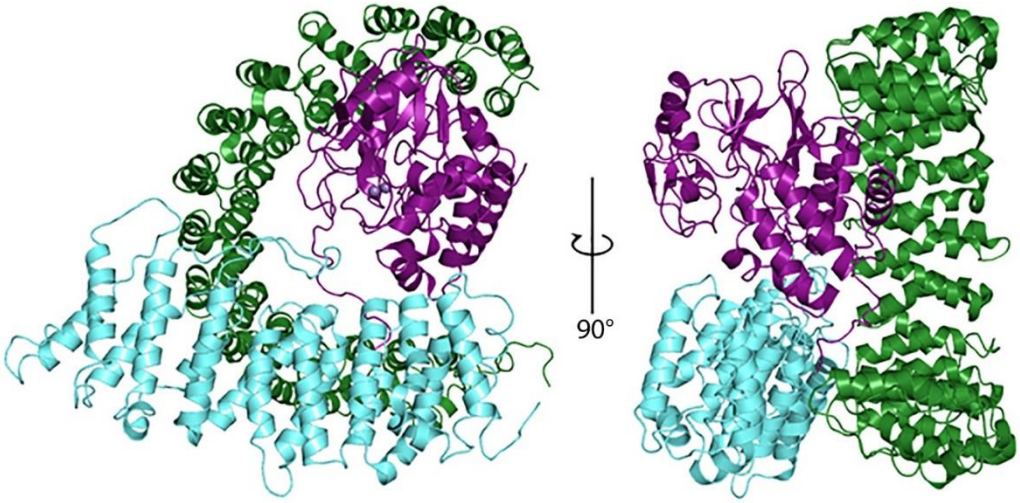


Fig 4.5: The crystal structures of the PP2A holoenzyme complex from humans. The complex is helix rich, consisting of the scaffold Aα subunit (green), the catalytic Cα subunit (purple) and the regulatory B56γ1 subunit (blue), which may combine in multiple isoforms. PDB code: 2IAE (Cho and Xu, 2007).

PP2A regulates a large cohort of proteins involved in various cellular processes through direct or indirect interactions including β-catenin from the Wnt signalling pathway (Eichhorn et al., 2009, Polakis, 2012), G1/S cyclins in cell cycle transition (McCourt et al., 2013), MEK and ERK of the MAPK/ERK pathway (Ciccone et al., 2015), AKT and GSK3β of the PI3K/AKT cascade (Andrabi et al., 2007, Janssens et al., 2008) as well as inhibition of PIM1 and Jak2 which inhibit the pro-apoptotic molecule BAD (Aho et al., 2004, Samanta et al., 2009, Oaks et al., 2013). A summary of these direct and indirect interactions is shown in Figure 4.6. Although PP2A is involved in regulation of a wide variety of cellular pathways, generally a negative trait when considering drug targets due to a high potential for causing unwanted side effects, the subunit specificity gives scope for high tissue specificity PP2A-activating drugs (PADs) as viable therapeutics. Some examples include *forkosil/1,9-dideoxyforkosil (Neviani et al., 2005)*, Fingolimod (Kappos et al., 2010, Ingwersen et al., 2012) and *OP449 (Agarwal et al., 2014)*. In addition to PADs, targeting inhibitors of PP2A also present possible drug targets. Of particular interest is the cancerous inhibitor of protein phosphatase 2A, which is the target of this study

4.2. Methods

Recombinant protein vectors were designed based on the initial bioinformatics work presented below, with two constructs produced encompassing the N-terminal residues 1-191, preceding the predicted transmembrane domain, and the C-terminal coiled coil domain (residues 635-882). Vectors were transformed into *E. coli* cells and purification attempted with a variety of chromatographic techniques.

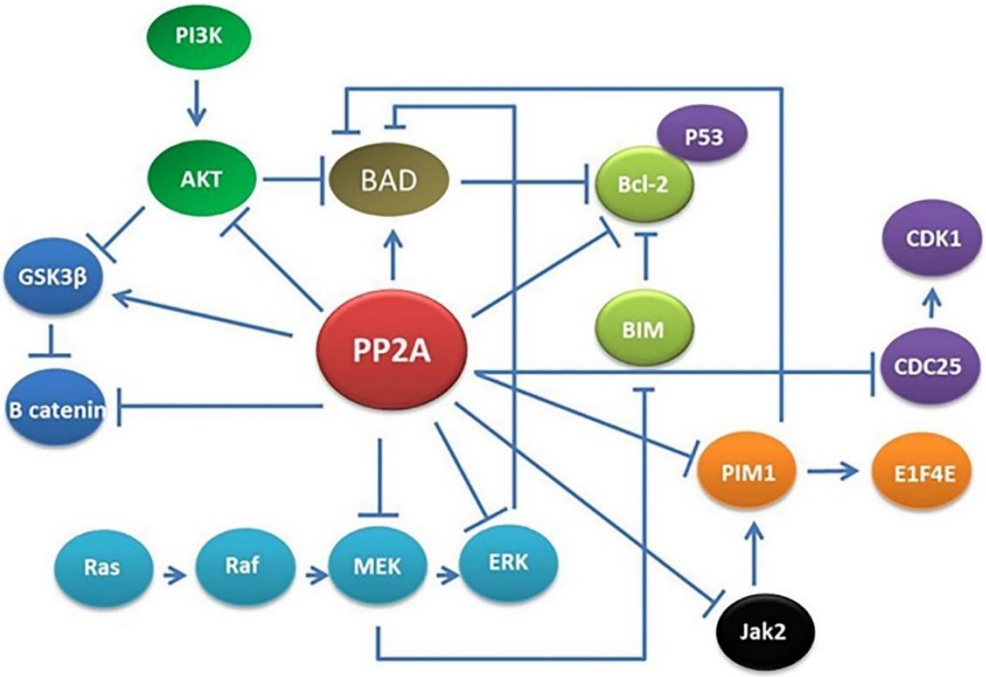


Fig. 4.6: The PP2A network of direct and indirect interaction including the MAP-K/ERK (light blue/light green), PI3K/AKT (dark green/dark blue), Jak2 (black/orange) and CDK (purple) pathways Ciccone et al. (2015).

4.2.1. Bioinformatics

The full length human CIP2A sequence was analysed with the basic local alignment search tool: protein (BLASTp) servers to find homologues and allow sequence alignment. The full length sequence was also with various secondary structure prediction servers including, InterProScan5 (Jones et al., 2014), Motif scan, PSIPRED (McGuffin et al., 2000) and JPRED (Cole et al., 2008). Function and disorder predictions were carried out on the full length sequence with FFPRED (Buchan et al., 2010) and DISOPRED, respectively. Crystallisation propensity of the full length sequence, half of N-terminal region (residues 1-191) and the C-terminal half of the protein (residues 635-882) was also assessed with PPCpred (Mizianty and Kurgan, 2011) and XtalPred (Slabinski et al., 2007).

4.2.2. Vectors

N and C terminal constructs for CIP2A were produced from residues 1-191 (CIP2A-N) and residues 635-882 (CIP2A-C). Sequences were cloned into pET-28 b (+) vectors between NdeI and XhoI restriction sites to produce a thrombin cleavable His_6 tag at the N-terminus and a non-cleavable His_6 tag at the C-terminus. Sequences were produced and cloned into pET-28 b (+) (Fig. 4.7) and Puc57 vectors by GenScript.

4.2.3. Expression Trials

CIP2A-N and CIP2A-C were transformed into Rosetta D2 and BL21 *E. Coli* cell in the presence of 30 µg/ml kanamycin by standard methods. Starter cultures were grown in 10 ml of LB broth with 30 µg/ml kanamycin at 37°C overnight. 100 ml of LB broth was inoculated with 1 ml of starter culture and grown at 37°C with shaking to an OD_{600} between 0.4-0.6; cells were then induced with 0.1 mM, 0.5 mM or 1.0 mM IPTG at 22°C. Pellets were harvested by centrifugation at 5000 rpm (Sorvall SLC6000 rotor) for 10 minutes at 4°C and resuspended in buffer A (0.1 M Tris-HCl pH 8.0, 0.5 M NaCl, 20 mM Imidazole, 2% glycerol) plus 10 mM $MgCl_2$, 10 µg/ml DNAse and 300 µg of lysozyme per gram of pellet. Cells were lysed by sonication in two cycles of 45 and 30 seconds and centrifuged for 15 minutes at 16,000 rpm (Sorvall SS34 rotor) at 4°C. Cells prior to induction and cleared lysate were loaded onto an SDS PAGE gel to assess expression levels.

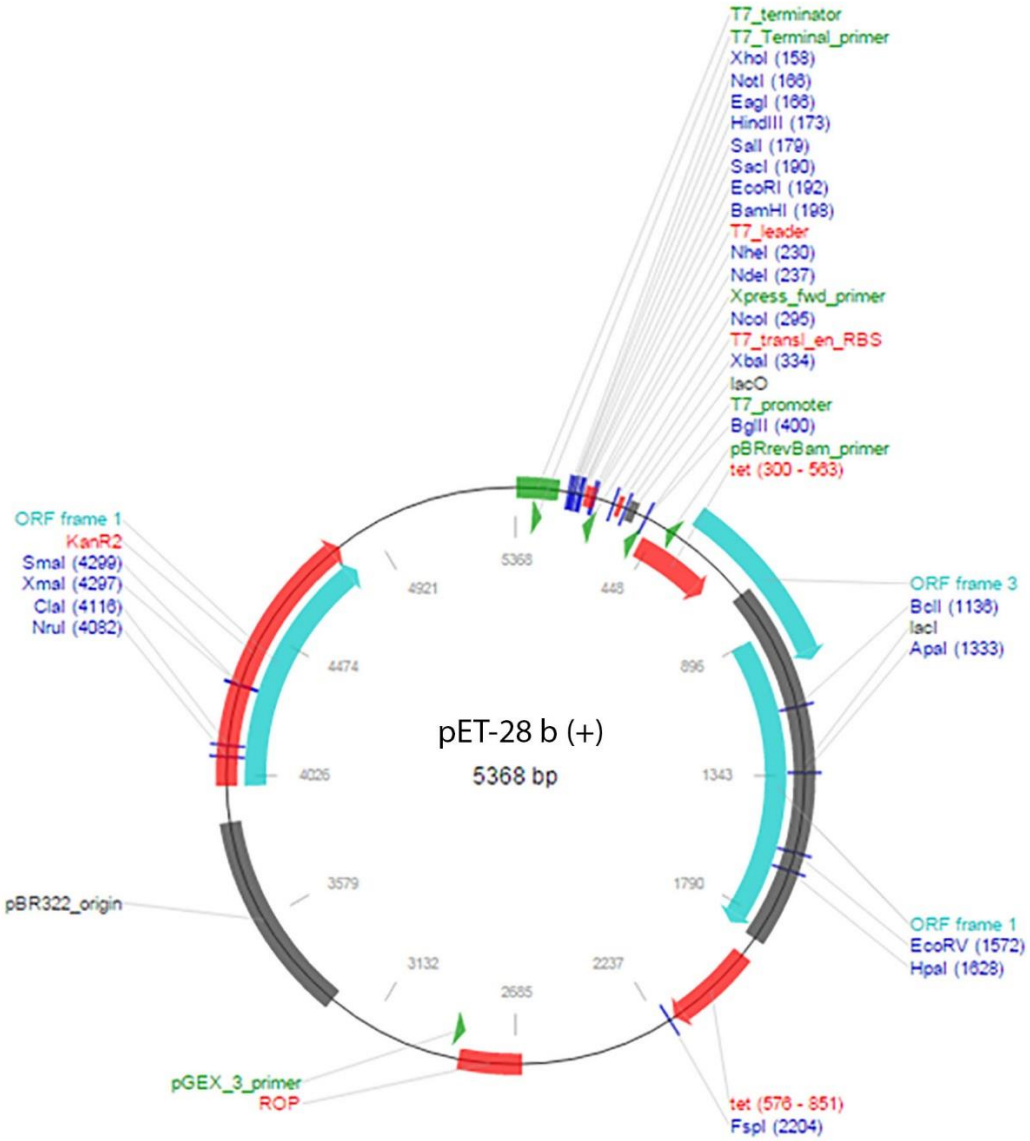


Fig. 4.7: pET-28 b (+) vector map. CIP2A N- and C-terminal constructs were inserted between XhoI and NdeI restriction sites to give N-terminal thrombin cleavable His_6 tag and non-cleavable C-terminal His_6 tag.

4.2.4. Protein Purification

Scaled up 1L LB cultures for CIP2A-N and CIP2A-C constructs were grown and harvested as detailed in the section above, with induction performed with 0.1 mM IPTG at 22°C. Clarified lysate was filtered with a 0.22 µm syringe filter and loaded onto a 5 ml Ni-NTA affinity column (GE Healthcare) pre-equilibrated in buffer A. Affinity chromatography was performed using an AKTA purifier (GE Healthcare) according to manufacturer's instructions. The Ni-NTA column was washed

with a linear gradient to 10 % buffer B (0.1 M Tris-HCl pH 8.0, 0.5 M NaCl, 1 mM TCEP, 20 mM Imidazole, 2% glycerol and 0.5 M imidazole) over one column volume (CV) followed by 5 CV of 10% buffer B. CIP2A-N and CIP2A-C were eluted with a linear gradient of increasing buffer B from 10-100% over 6 CV. CIP2A-C fractions were pooled to a final volume of 6 ml and 3 ml was loaded onto either Superdex 75 16/60 or Superdex 200 10/300 gel filtration columns equilibrated in size exclusion chromatography (SEC) buffer (20 mM Tris-HCl, pH 8, 300 mM NaCl, 1mM TCEP).

4.3. Results

4.3.1. Bioinformatics

4.3.1.1. BLASTp

BLASTp of the full length human CIP2A sequence reveals no homologues to CIP2A other than CIP2A from other species and proteins of unknown function. CIP2A appears well conserved between species, as shown in Figure 4.A1, with the CIP2A sequence from chimpanzees, Rhesus macaques, white-tuffed-ear marmosets, black flying foxes, naked mole rats and mice being 99.4% conserved at best and 87.2% conserved at worst.

4.3.1.2. PSIPRED, FFPred and DISOPRED

PSIPRED analysis (McGuffin et al., 2000) predicts CIP2A to be completely helical, with possible disordered protein binding regions at the N and C-terminal ends of the sequence (Fig. 4.A2). A disordered region is also predicted between residues 551-596. FFPred corroborates the largely helical structure predicted by PSIPRED, but suggests the N-terminal region forms an ordered helical structure while the C-terminal helices are more disordered. A disordered coil is also predicted between residues 562-597, similar to PSIPRED (Fig. 4.A3). In addition to secondary structure predictions FFPRED identifies a number of potential phosphorylation sites (red boxes in Fig. 4. A3)

and glycosylation sites (green boxes in Fig. 4.A3). DISOPRED suggests a similar ordered N-terminal arrangement and disordered C-terminus.

4.3.1.3. InterProScan5 and Motif Scan

The secondary structure predictions reported above are in line with the domain prediction from InterProScan5 and Motif Scan predictions (Fig. 4.8). Domain analysis predicts armadillo domains in the N-terminal half of CIP2A (residues 48-309), three helix repeats which form an ordered right handed super helical tertiary structure (Conti and Kuriyan, 2000, Conti et al., 1998, Daniels et al., 2001). A leucine zipper (residues 331-352) and transmembrane region (residues 192-212) are also predicted in and near the predicted armadillo repeats, respectively. Finally, a coiled coil region (residues 635 and 880) is predicted, which are regularly found to be intrinsically unstructured (Mason and Arndt, 2004, Chana et al., 2002) to facilitate the wide variety of protein-protein interactions and other functions associated with these domains (Burkhard et al., 2001).

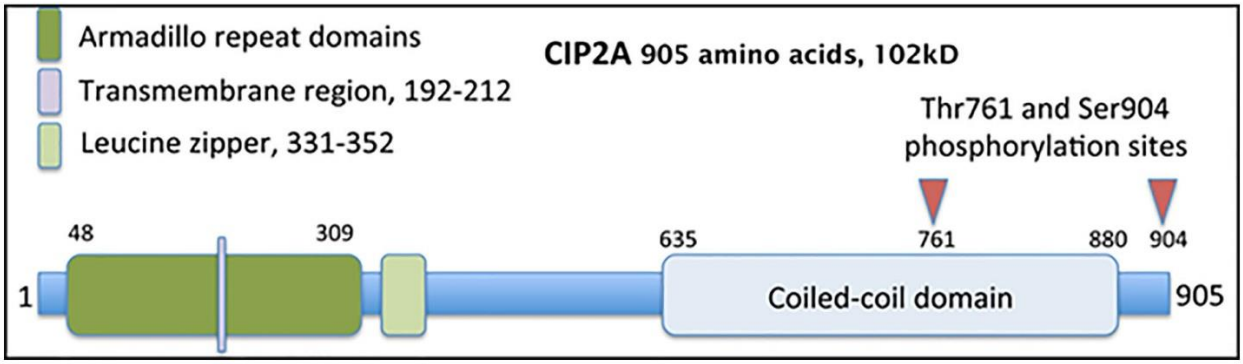


Fig. 5.8: Predicted domains of full length human CIP2A based on preductions by InterProScan5 and Motif Scan with experimentally determined phosphorylation sites.

4.3.1.4. Crystallisation Prediction

PPCpred predictions (Mizianty and Kurgan, 2011) for the full length CIP2A suggests a very high propensity for production, purification and crystallisation to fail, with probabilities of 62%, 52% and 84% respectively.

Predictions for two constructs of the coiled coil region of CIP2A, 636-833 and 619-890, both showed a lower propensity for failure in production, purification and crystallisation. However, the construct for residues 619-890 presented with a much higher propensity for crystallisation to fail (47%) compared to residues 635-882 (19%). These results were corroborated by XtalPred (Slabinski et al., 2007) which suggested both full length and coiled coil domain alone would be difficult cases based on the large regions of disorder.

PPCpred analysis of the N-terminal sequence used to generate the model in Figure 4.9 predicted a high probability of failure at protein production (51%), purification (55%) and crystallisation (42%). As with the C-terminal region, XtalPred predicted this N-terminal construct to be a difficult target. A shorter N-terminal construct (residues 1-191) was also tested based on predictions of a transmembrane domain at residues 192-212. The shorter N-terminal construct shows a much better propensity to protein production and crystallisation with failure predicted at 35% and 24%, respectively. However, the chance of failure in purification was still very high at 66%. Again XtalPred predicted this to be a difficult target based on a high instability index. Despite a high prediction of failure CIP2A was pursued as a challenging target as it is a high value target which could be potentially very rewarding for future drug discovery. It should also be kept in mind that crystallisation prediction is largely based on previously solved structures and CIP2A has few homologues based on sequence and only a small number of secondary structure homologues for the N-terminal half of the protein.

4.3.1.5. Structure and Function Prediction

Building on earlier bioinformatics work in the Molecular Biophysics Group at the University of Liverpool by third year hons student Sam Pringle, new insights into the predicted structure and function of CIP2A have been obtained.

BLASTp (Altschul et al., 1990) of full length CIP2A produces various coiled coil structures relating to residues 500-905 of CIP2A (CIP2A-CT), the N-terminal half (1-500) however, produces no homologous sequences. The presence of a C-terminal coiled coil was confirmed by MultiCoil2 (Trigg et al., 2011) and logiCoil (Vincent et al., 2013), with three potential dimeric coiled coil regions. As with many coiled coil domains CIP2A-CT was predicted to be intrinsically unstructured by IUpred (Dosztanyi et al., 2005), which also predicted four possible ligand binding sites. In addition to being intrinsically unstructured CIP2A-CT is predicted to have a high propensity for self-aggregation, as judged by AmylPred2 (Tsolis et al., 2013). These properties may explain the difficulties faced during purification, as discussed later.

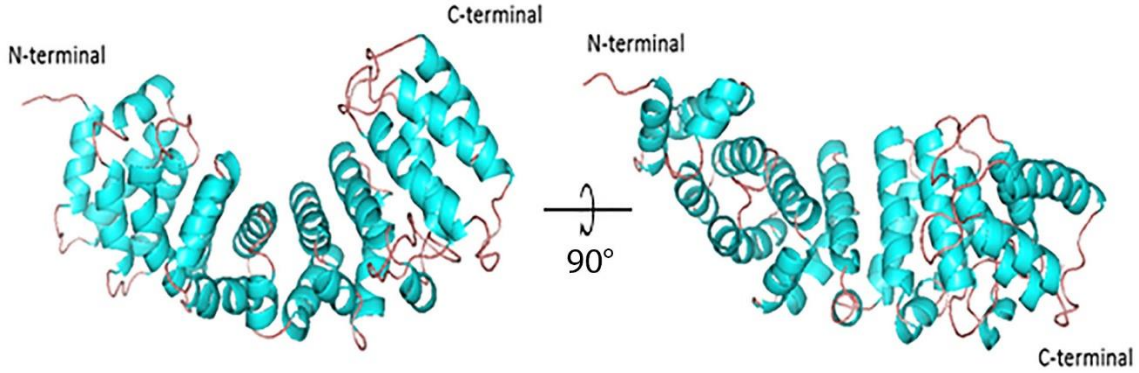


Fig. 4.9: Predicted CIP2A N-terminus model generated by HHpred modeller from CIP2A residues 50-420 and several secondary structure homologues. Helices are coloured blue with red connecting loops.

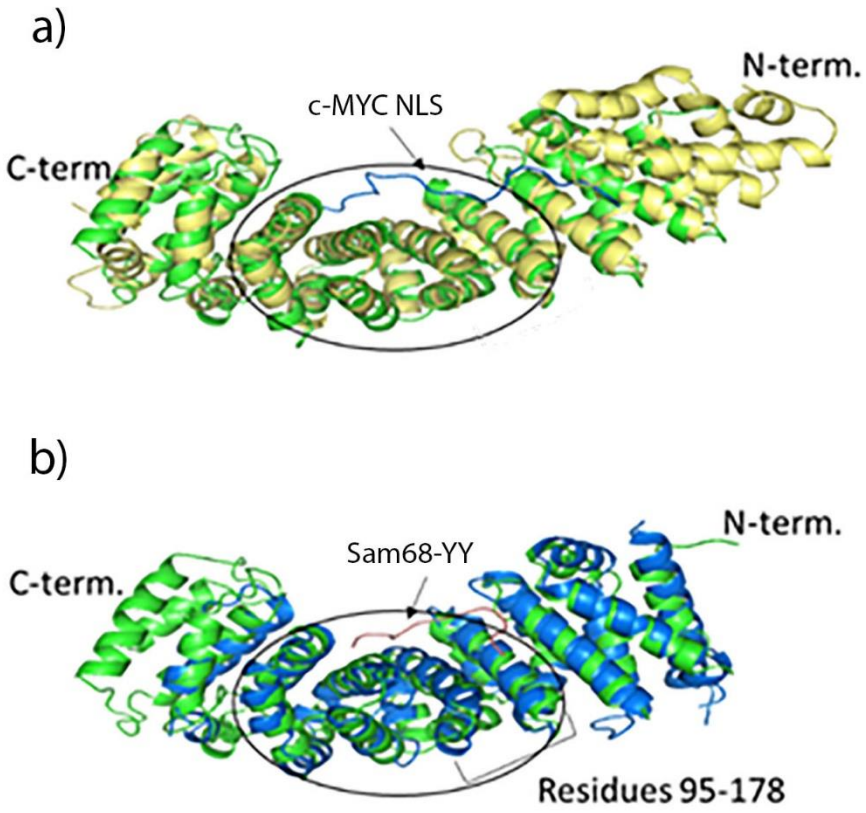


Fig. 4.10: Ligand binding analysis of CIP2A-NT. a) CIP2A-NT (green) aligned against c-MYC NLS peptide bound Karyopherin α (yellow). b) CIP2A-NT (green) aligned against Sam68-YY domain bound APC. The conserved central region of the predicted model is circled in black.

Although the N-terminal domain showed no sequence identity to other proteins during BLASTp searches, analysis with HHpred (Hildebrand et al., 2009) found secondary structure homologues for residues 40-420. These included the importin α-1 subunit (98.3%), catenin β-1 (98.1%) and karyopherin β2 (95.7%); despite a sequence homology of ≤ 20% at best. All secondary structure homologues have Armadillo (Arm) repeats, three helix repeats that form right handed super-helical structures (Coates, 2003, Huber et al., 1997), usually involved in protein-protein interactions (Conti and Kuriyan, 2000, Conti et al., 1998, Daniels et al., 2001). InterProScan5 (Jones et al., 2014) corroborated the presence of potential Arm repeats between residues 50-420 of CIP2A.

A model of N-terminal CIP2A (CIP2A-NT) was generated from secondary structure homologues (PDB codes: 4HXT, 4B8J, 4EV8, 4FDD, 43GA and 2OF3) using protein threading with HHpred modeller (Fig. 4.9) (Hildebrand et al., 2009, Söding et al., 2005) . CIP2A-NT was validated by VERIFY3D (Eisenberg et al., 1997), QMEAN6 (Benkert et al., 2011) and Ramachandran plots (Lovell et al., 2003), and assessed for conserved structural elements with ConSurf (Goldenberg et al., 2009). Six homologous structures were found for CIP2A-NT, four of which were bound to nuclear localisation sequences (NLS). Superposition of Karyopherin α bound to the c-Myc NLS peptide (Conti and Kuriyan, 2000) and adenomatous polyposis coli (APC) bound to the Sam68 NLS peptide (Morishita et al., 2011) with CIP2A-NT showed binding relative to CIP2A-NT's highly conserved central region (Fig. 4.10a-b). Homology with Karyopherin α bound to the c-Myc NLS peptide is particularly interesting as CIP2A is known to interact directly with c-Myc (Junttila et al., 2007).

4.3.2. Expression Trials

Expression trials for CIP2A-N and CIP2A-C showed clear over expression in Rosetta D2 (Fig. 4.19) and in BL21 cells (data not shown). Solubilised whole cells in lanes 2 and 7 of Figure 4.11 for CIP2A-C and CIP2A-N respectively, show very little leakage of the plasmid before induction, with a clear increase in expression following induction at ~ 22 kDa for CIP2A-N and ~ 35 kDa for CIP2A-C, in line with the predicted molecular weights based on sequence (21 kDa and 30.7 kDa, respectively). Increasing the concentration of IPTG on both constructs appears to have little effect on the overall levels of expression.

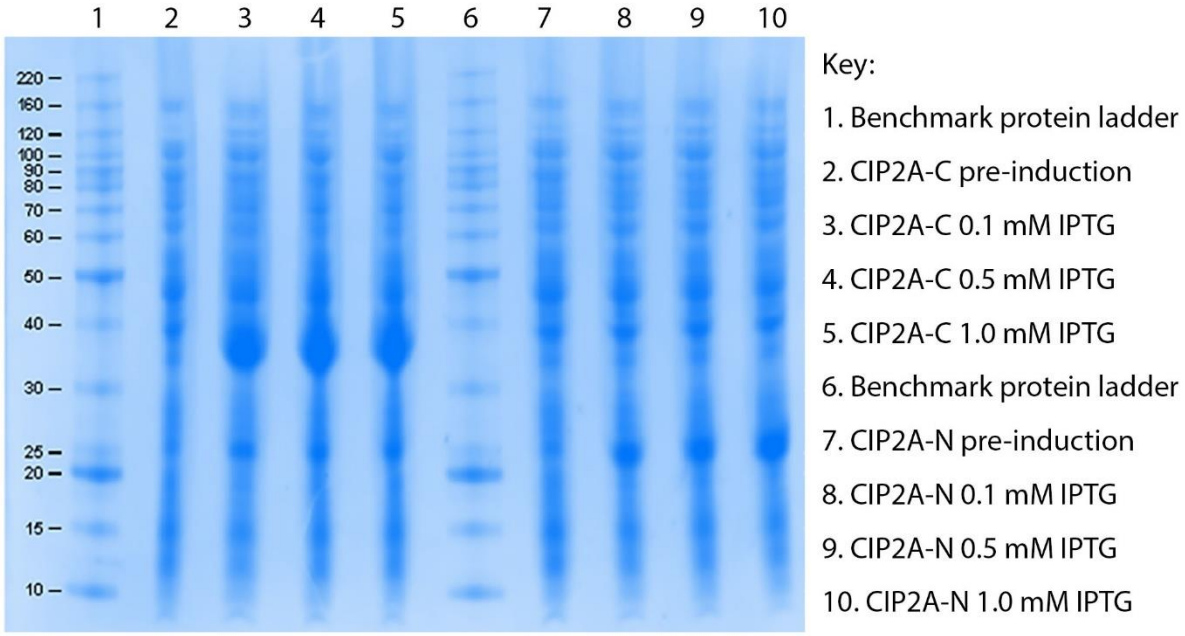


Fig. 4.11: SDS-PAGE of CIP2A-N (22 kDa) and CIP2A-C (34 kDa) expression trials in Rosetta D2 *E. coli* for increasing IPTG concentrations.

4.3.3. Protein Purification

Purification of CIP2A-N by nickel affinity chrmoatography gave a large peak of 350 mAU during the wash with 10% buffer B (Fig. 4.13a), which consisted largely of impurities and some CIP2A-N protein as seen in lane 3 of Figure 4.12. This intial larger peak is followed by two smaller peaks, labelled peak two and three in Figure 4.13. Peak two elutes at around 225 mM imidazole with a more prominent band around 22 KDa with many impurities (Fig. 4.12, lane 4). Peak three elutes around 300 mM imidazole with very high purity CIP2A-N (Fig. 4.12, lane 3), however the absorbance of 150 mAU over ~ 6 ml gives only 1 mg of almost pure protein from 6 litres of cells. Repeats of CIP2A-N protein production with increased IPTG concentration on induction saw the formation of inclusion bodies and significantly reduced yield (data not shown).

CIP2A-C nickel affinity chromatography gave a single large peak of 1100 mAU spread over 160-500 mM imidazole, with maximum absorbance around 300 mM imidazole (Fig. 4.13b). SDS-PAGE of samples from the peak show a large number of impurities in the left hand side and centre of

the peak (Fig. 4.12, lanes 7-8) with more pure CIP2A-C on the right hand side of the peak (Fig. 4.12, Lane 9).

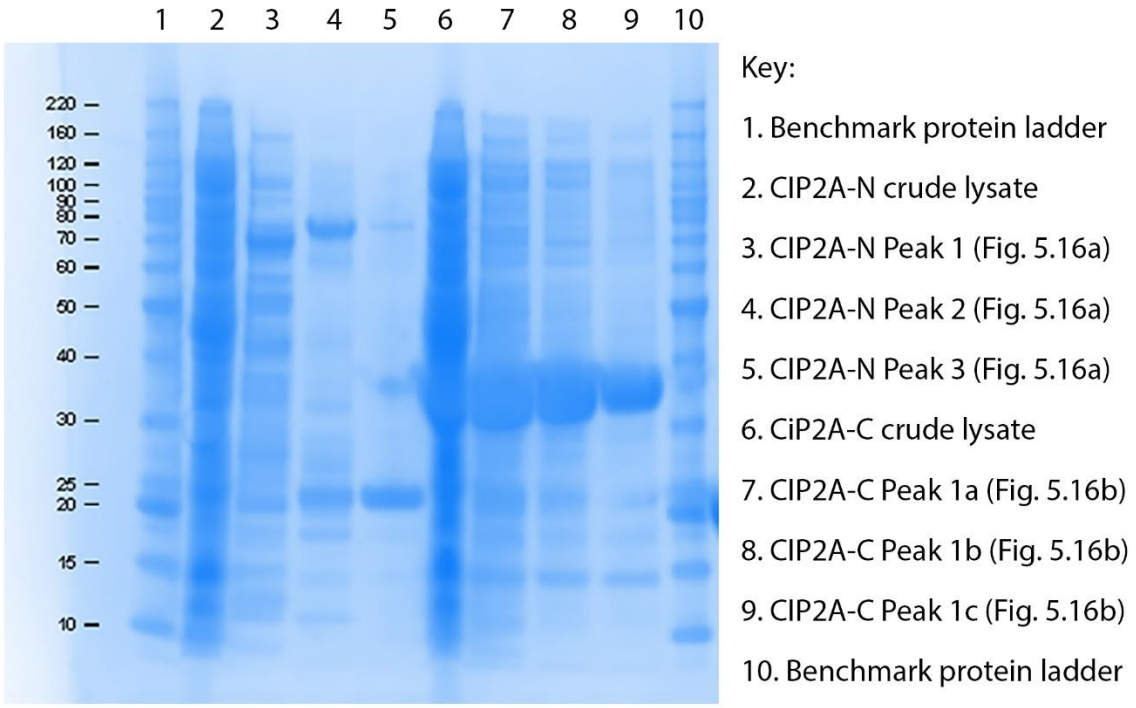


Fig. 4.12: SDS-PAGE of CIP2A-N (left) and CIP2A-C (right) Ni-NTA purification.

Dialysis of CIP2A-N and CIP2A-C to remove excess imidazole for gel filtration caused both proteins to crash out of solution. Addition of 1-3 mM DTT to CIP2A-C prevented precipitation during dialysis as shown in Figure 4.14, where samples of CIP2A-C post dialysis (lanes 3-5) and post centrifugation to remove any precipitate (lane 6-8) show no difference in the SDS-PAGE profile. However, addition of DTT to binding buffer caused nickel to leach from the column so it was replaced with TCEP in the final binding buffer. Addtion of DTT and TCEP to CIP2A-N had no effect on precipitation during dialysis, however, buffer exchange by concentration proved more effective, but still presented with some loss of protein. Reduction in CIP2A-N was too severe to continue but enough CIP2A-C remained for gel filtration. Pooled CIP2A-C eluates ran down a Superdex 75 16/60 and Superdex 200 10/300 gel filtration columns both eluted in the void volume (Fig. 4.16a and b,

respectively), suggesting aggregation of CIP2A-C with itself and the other impurities seen in SDS-PAGE analysis (Fig. 4.15).

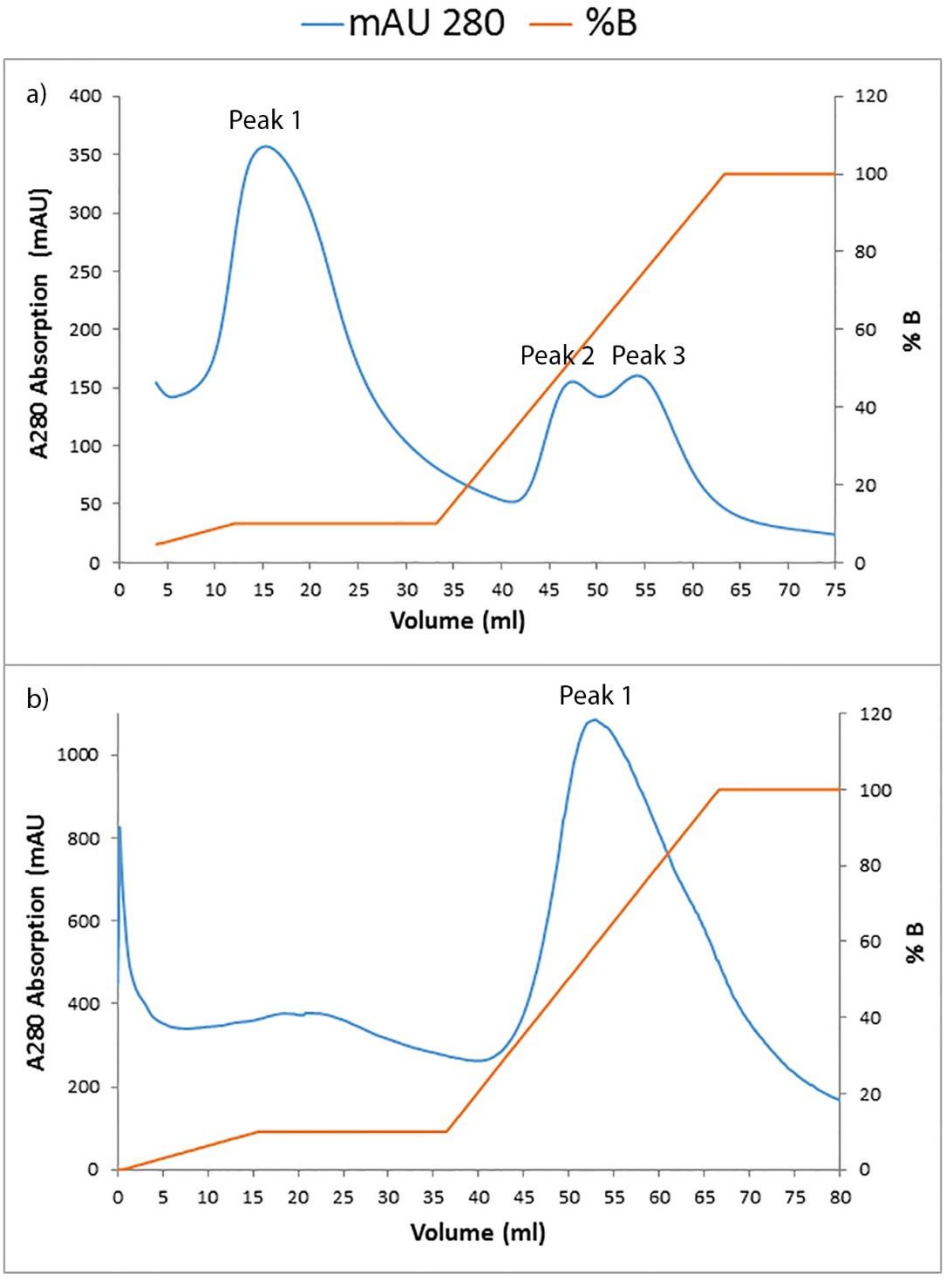
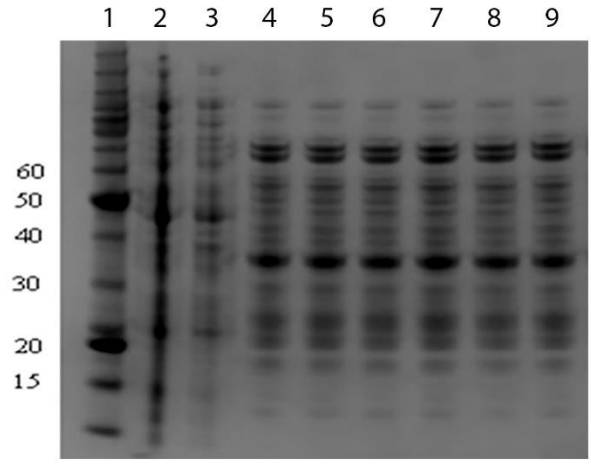
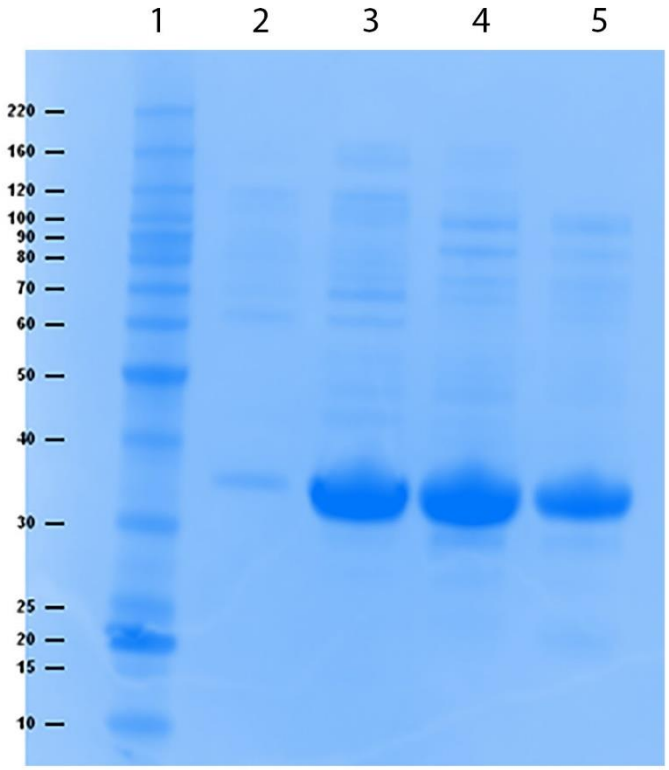


Fig. 4.13: CIP2A nickle affinity A280 elution peaks (blue) and perfectage of buffer B (orange) plotted against volume. a) CIP2A-N purification. b) CIP2A-C purification.



Key:
1. Benchmark protein ladder
2. CIP2A-C Crude Lysate
3. CIP2A-C Ni-NTA flow through
4. CIP2A-C with 1 mM DTT post dialysis
5. CIP2A-C with 2 mM DTT post dialysis
6. CIP2A-C with 3 mM DTT post dialysis
7. CIP2A-C with 1 mM DTT post centrifugation
8. CIP2A-C with 2 mM DTT post centrifugation
9. CIP2A-C with 3 mM DTT post centrifugation

Fig. 4.14: SDS-PAGE of CIP2A-C DTT solubility tests. CIP2A-C samples before dialysis, after dialysis and after centrifugation in the presence of increasing DTT.



Key:

1. Benchmark protein ladder
2. Peak 1a (Fig. 4.16a)
3. Peak 1b (Fig. 4.16a)
4. Peak 1c (Fig. 4.16a)
5. Peak 1d (Fig. 4.16a)

Fig 4.15: SDS-PAGE of CIP2A-C size exclusion chromatography with a Superdex 75 16/60 gel filtration column.

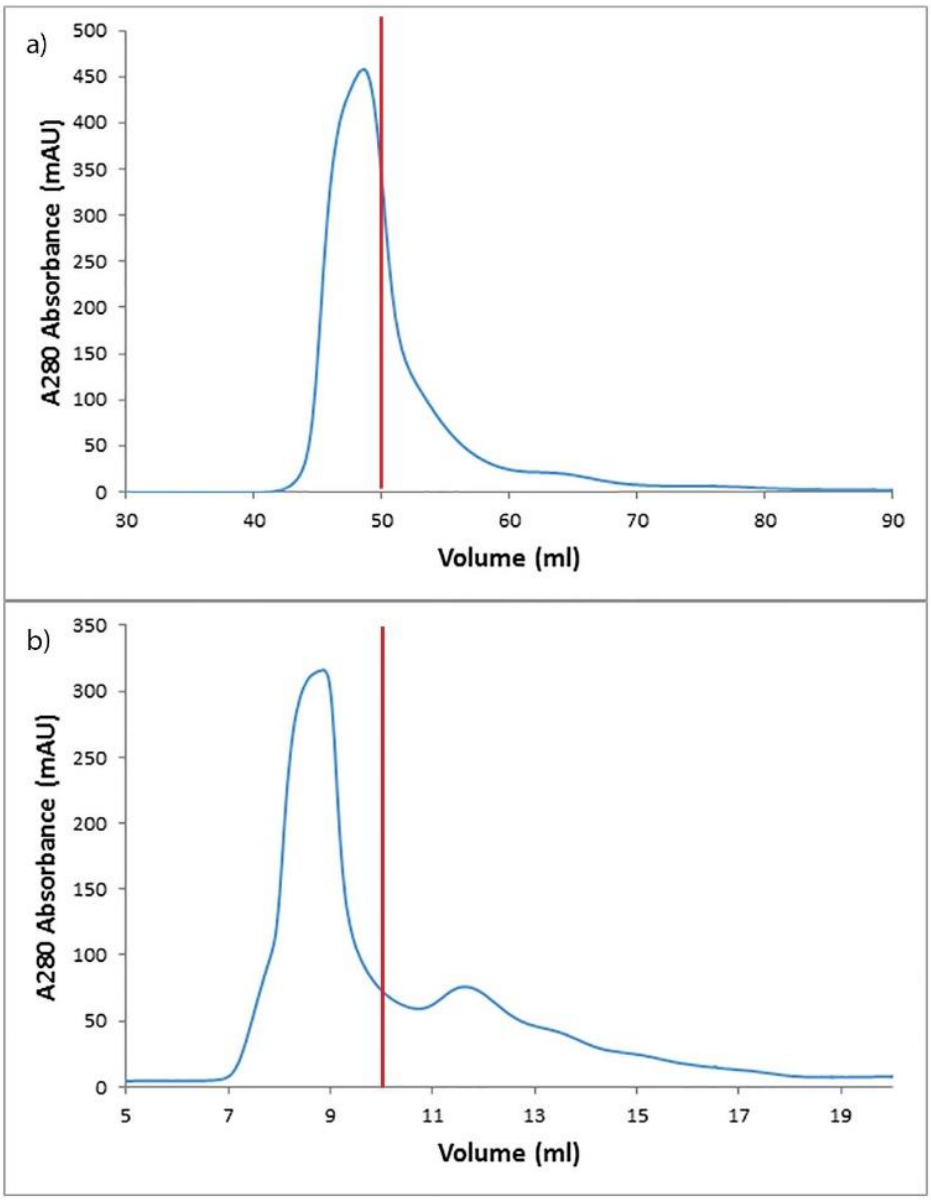


Fig. 4.16: ClP2A-C size exclusion chromatography peaks at A280 against elution volumn. Red lines represent the void volume. a) Gel filtration with a Stuperdex 75 16/60 column. b) Gel filtration of the same sample with a superdex 200 10/300 column.

4.4. Discussion

4.4.1. Bioinformatics, Expression Trails and Purification

Expression of both CIP2A-N and CIP2A-C constructs in *E. coli* cells was successful, although initial expression of CIP2A-N was low and attempts to increase yield saw the formation of inclusion

bodies and a further reduction in yield. CIP2A-C purification was more successful in terms of yield but purity was poor following Ni-affinity purification with a tendency for protein to aggregate in preparation for and during gel filtration. Recloning CIP2A-C into a vector with a single cleavable his tag may allow elution over a narrower imidazole concentration and improve purity, as well as permitting a second round of Ni-affinity purification following tag cleavage. Alternatively addition of a solubility tag such as small ubiquitin-related modifier (SUMO), glutathione S-transferase (GST) or maltose-binding protein (MBP) may help prevent aggregation during purification (Walls and Loughran, 2011). The large number of post translational modifications identified by FFPRED may also indicates a need for expression in mammalian or insect cells to allow proper folding.

In light of more in-depth bioinformatics analysis following initial attempts at purification, a redesign of the N-terminal construct to encompass the entire armadillo repeat may be beneficial. However, the high self-aggregation propensity of the N-terminal domain predicted by AmylPred2 suggests this domain may still prove a difficult case for expression, purification and crystallisation. As with CIP2A-C, the large number of phosphorylation and glycosylation sites predicted by FFPRED may contribute to the aggregation of the N-terminal construct. Recloning into mammalian or insect cell vectors may be beneficial in preventing aggregation and improving the yield of protein by allowing post translational modifications for more effective folding. However, all structural homologues of the full length N-terminal region identified via ConSurf (Goldenberg et al., 2009) were expressed and purified using *E. Coli* cells (Ishiyama et al., 2010, Chang et al., 2012, Chang et al., 2014, Conti and Kuriyan, 2000); so expression in insect or mammalian cells may not be necessary with the full Arm-domain.

4.5. Conclusions

CIP2A was identified as both a biologically and medically important target for future drug discovery approaches and also as a suitable difficult and challenging candidate for eventual sulphur

phasing experiments on I23. However, expression and purification to the levels required for crystallisation proved too challenging during the time scale of this thesis. Clearly, it is necessary that considerable time is allocated to improving yield and purity as well as to examining alternative molecular biology approaches to those taken here. Work on producing and crystallising full length CIP2A is continuing within the Molecular Biophysics Group in partnership with Prof. Richard Clark from the Institute of Translational Medicine at the University of Liverpool.

4.6. Appendix

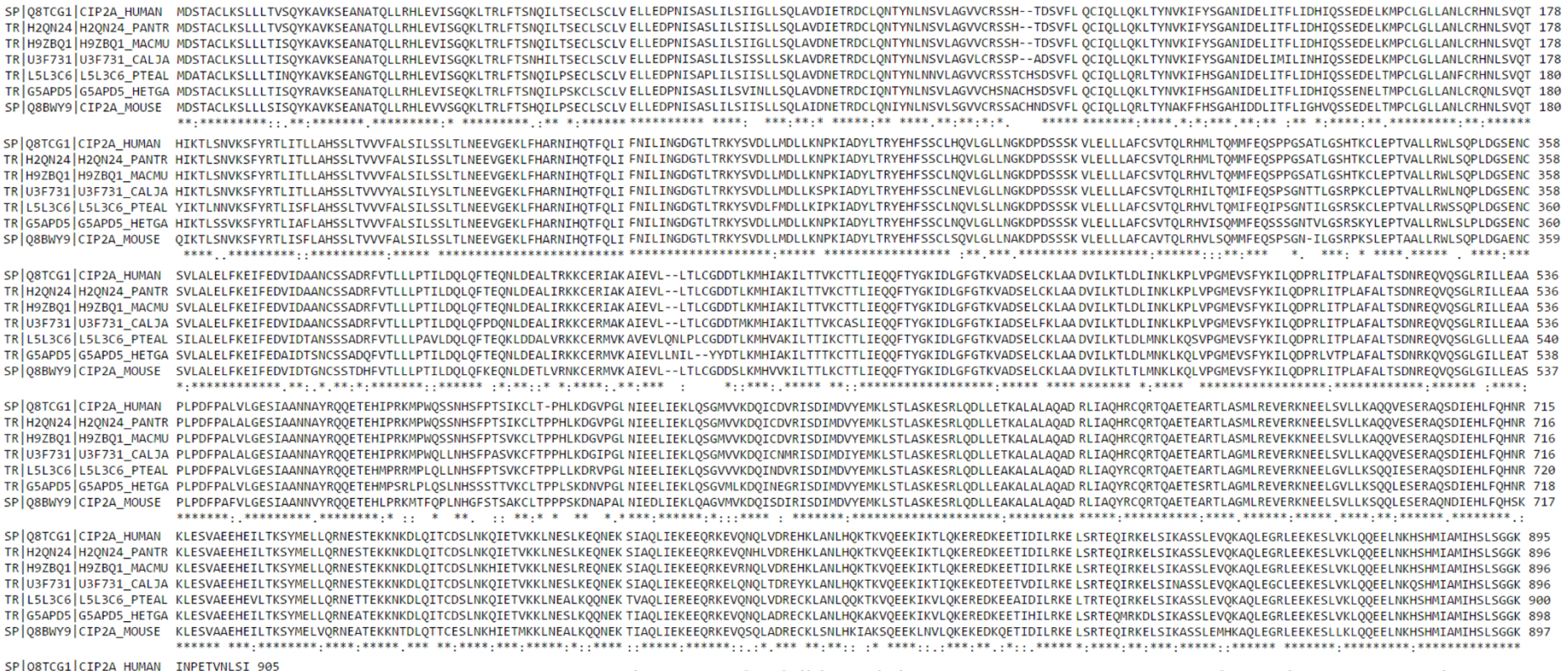


Fig. 4.A1: BLASTp sequence alignment for full length human CIP2A against CIP2A from chimpanzees, rhesus macaques, white-tuffed-ear marmosets, black flying foxes, naked mole rats and mice.

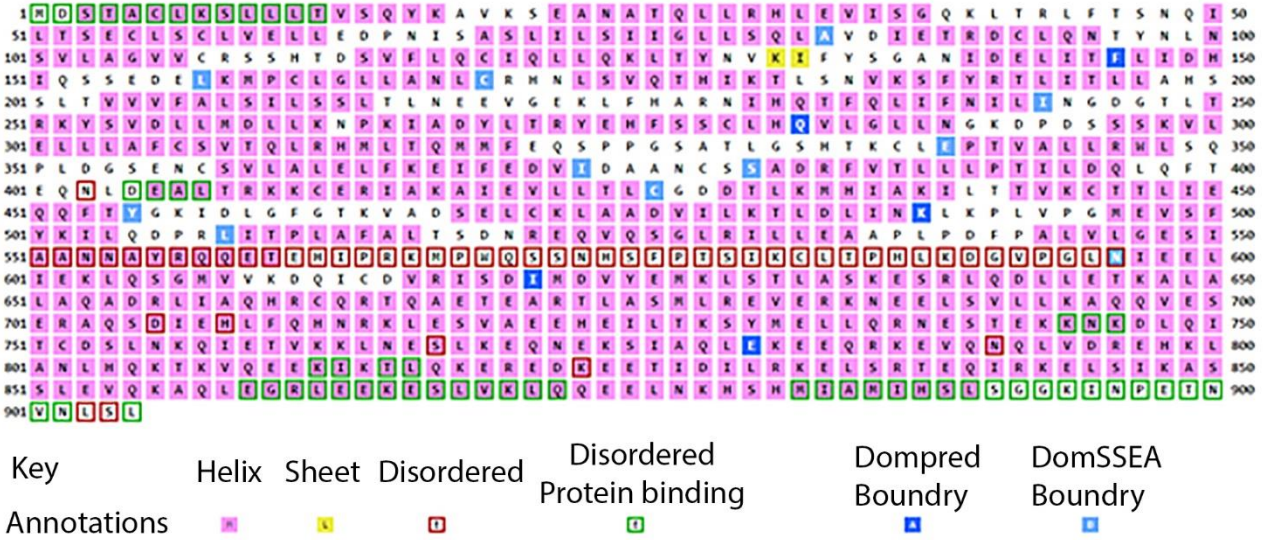


Fig. 4.A2: PSIPRED secondary structure predictions for full legnth Human CIP2A.

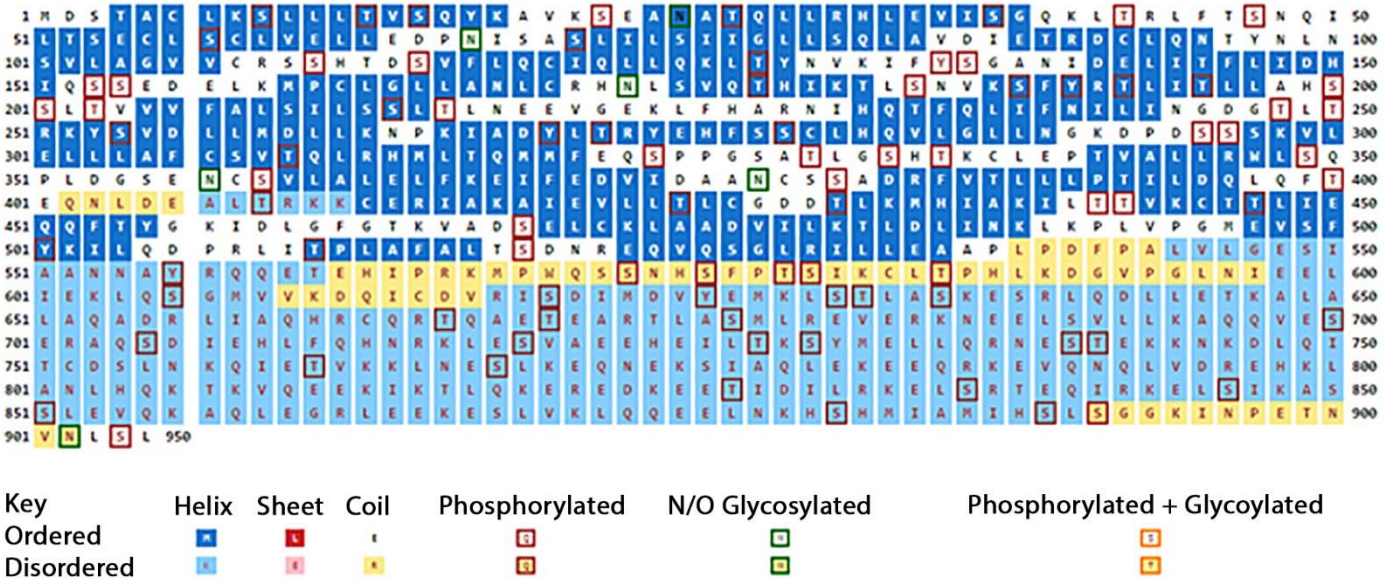
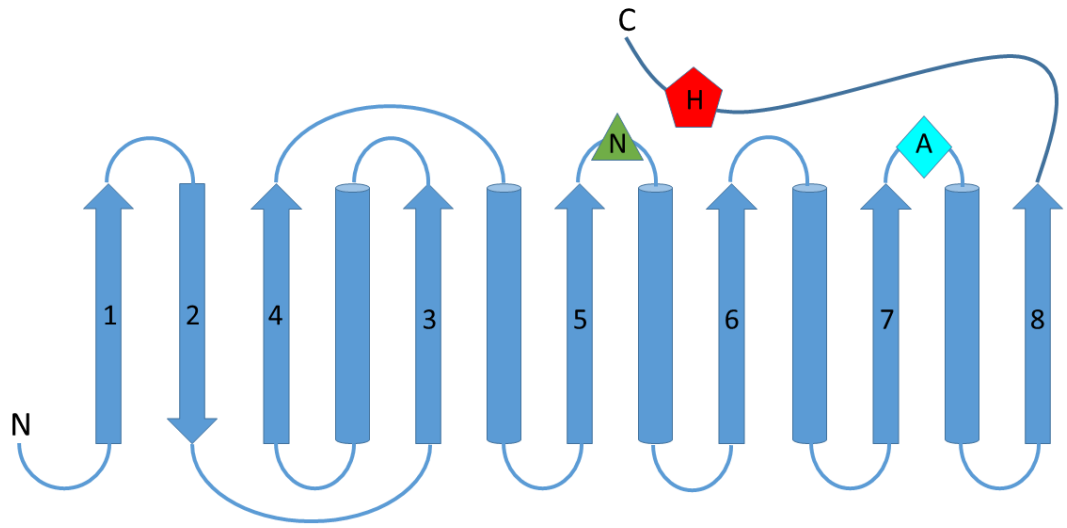


Fig. 4.A3: FFPred secondary structure predictions for full length human CIP2A.

Chapter 5 – *Achromobacter xylosoxidans*
Targets for S-SAD phasing



5.1. Introduction

In parallel with the unsuccessful attempts to purify and crystallise fragments of CIP2A for phasing experiments on I23 at Diamond Light Source (Chapter 5), a set of more easily attainable and, according to a review of the relevant literature, more soluble bacterial proteins were selected from the *Achromobacter xylosoxidans* (*Ax*) genome (unpublished data from Hasnain group at Liverpool) for expression, purification and crystallisation. Two protein targets were finally selected from a group of sixteen *Ax* proteins, initially selected on the basis of the ease of crystallisation, as judged by the number of homologues present in the PDB, and by the sulphur content of the protein; the aim being to provde relatively simple and difficult phasing cases for experiments using I23. The proteins selected were UTP-glucose-1-phosphate uridyltransferase (*Ax*UDT) and α/β hydrolase (*Ax*ABH). In addition to this, *A. xylosoxidans* is an opportunistic pathogen known to infect cyctic fibrosis sufferers (Cystic Fibrosis Foundation Patient Registry. Annual Data Report 2013. Cystic Fibrosis foundation, Bethesda, MD) and silencing of the equivalent genes encoding *Ax*UDT and *Ax*ABH in *P. eruginosa* were found to be essential to the survival of the bacteria (Winstanley et al., 2009); making both proteins potential targets for structure based drug design.

*Ax*UDT, also known as a UDP-glucose phosphorylase or UGPase, catalyses the addition of a UTP molecule to glucose-1-phosphate to produce UDP-glucose, a key component in the synthesis of glycogen (Alonso et al., 1995) for the production of glycolipids, glycoproteins and proteoglycans (Thoden and Holden, 2007). UGPases play an essential role in both prokaryotic and eukaryotic organisms but carry out the same reaction using completely unrelated proteins in terms of sequence and 3D-structure (Mollerach and Garcia, 2000, Mollerach et al., 1998). Owing to their role in the production of various virulence factors, UGPases have become important drug targets for use against Gram-negative bacteria such as *Escherichia coli* and *Streptococcus pneumoniae* (Genevaux et al., 1999, Bonofiglio et al., 2005). *Ax*UDT is a 276 residue 30 kDa protein with 10 sulphur atoms (three cysteines and seven methionines), giving a Bijvoet ratio of 0.32% at λ = 1 Å, 1.23% at λ = 2 Å,

2.47% at λ = 3 Å and 3.97% at λ = 4 Å, assuming one molecule in the asymmetric unit . Based on the Wang limit of 0.6% (Yao et al., 2005) this should be a relatively straightforward case for sulphur phasing at 'standard' beamline wavelengths.

*Ax*ABH is a member of one of the largest groups of structurally related enzymes which conveys an impressive cross-section of catalytic functions including but not limited to acetylcholinesterases, dienelactone hydrolases, lipases, thioesterases, serine carboxypeptidases, proline iminopeptidases, proline oligopeptidases, haloalkane dehalogenases, haloperoxidases, epoxide hydrolases, and hydroxynitrile lyases (Holmquist, 2000). α/β hydrolases have a well conserved fold of 8 to 11 parallel β-sheets flanked by α-helices, which houses the conserved catalytic triad (Nucelophile-Histidine-Acid) for hydrolysis as shown in Figure 5.1 (Marchot and Chatonnet, 2012). Additional highly variable motifs decorate the conserved α/β fold to form the active site and convey the wide variety of enzymatic functions. The exact enzymatic activity of *Ax*ABH is currently unknown. *Ax*ABH is a 311 residue 34.5 kDa protein with three sulphur atoms (three methionines, no disulphide bridges), giving a Bijvoet ratio of 0.16% at 1 λ = 1 Å, 0.62% at λ = 2 Å, 1.29% at λ = 3 Å and 2.06% at λ = 4 Å, assuming one molecule in the asymmetric unit . This is likely to be a very difficult case for S-SAD phasing using the wavelengths attainable at standard MX beamlines but represents a good test for S-SAD experiments at the long wavelengths expected for I23.

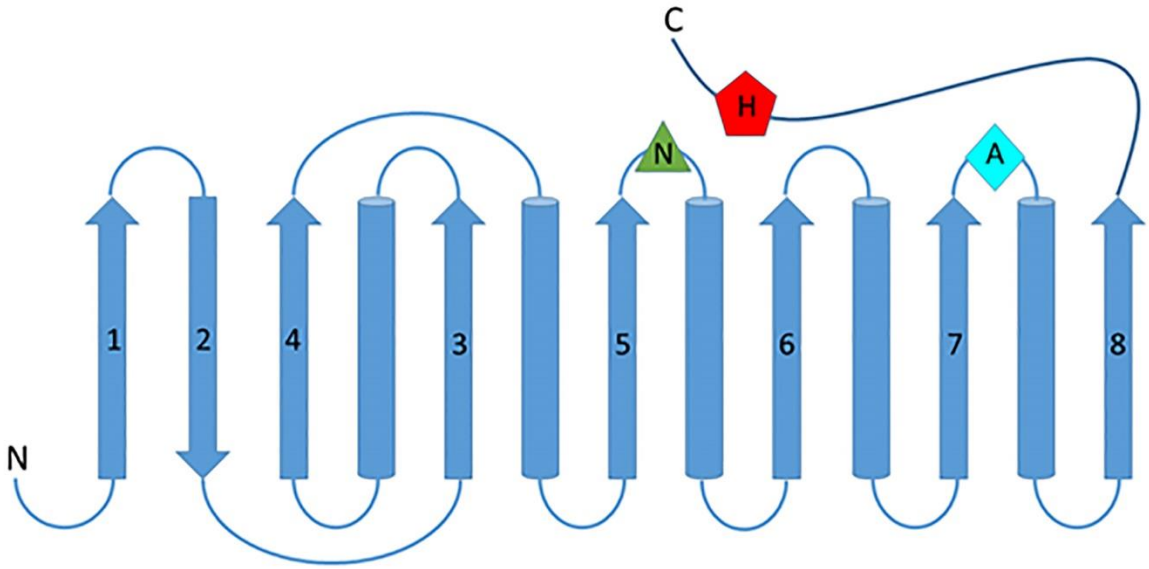


Fig. 5.1: A schematic diagram of the conserved α/β-hydrolase fold with β-sheets numbered from the N-C terminus as blue arrows and α-helices as blue cylinders. The position of the catalytic nucleophile-histamine-acid triad are shown as a green triangle, red pentagon and cyan diamond, respectively.

Both proteins were successfully expressed in *E. coli*, purified and crystallised. The crystallisation conditions were optimised in an attempt to produce high symmetry crystals, allowing for high redundancy anomalous data to be collected with minimal effects from radiation damage. This resulted in several crystal forms being available for study. However, owing to technical difficulties in the development of the I23 experimental end station, the anticipated long wavelength diffraction data for *Ax*UDT and *Ax*ABH crystals could not be collected. As a result phasing attempts were performed on *Ax*ABH and *Ax*UDT using molecular replacement and S-SAD and heavy atom soaking, respectively. After multiple phasing attempts involving several experiments, the structure of *Ax*UDT remained unsolved. *Ax*ABH was successfully phased by molecular replacement for several crystal forms, which have been studied in an attempt to reveal details of the hydrolytic process it catalyses.

5.2. Methods

5.2.1. Cloning, Expression and Purification

The Sequences for *Achromobacter xylosoxidans* UTP-glucose-1-phosphate uridyltransferase (*Ax*UDT) and α/β hydrolase (*Ax*ABH) were cloned into pET47b(+) (Fig. 5.2) between BamHI and HindIII to generate an N-terminal 3C protease cleavable His_6 tagged protein. Both plasmids were made by Genscript. Plasmids were transformed into BL21 *E. coli* cells in the presence of 30 mg/ml kanamycin. Small scale expression trials in 100 ml of LB media were performed with whole cells and cleared lysate ran on an SDS PAGE to assess expression levels. Expression was then scaled up to 2 litre cultures. Cultures were grown to an OD_{600} of 0.4-0.6 at 37°C and induced with 0.15 mM IPTG at 18°C for 16 hours. Cultures were spun at 5000 RPM (Sorvall SLC6000) for 10 minutes at 4°C and resuspended in buffer A (0.1 M Tris-HCl pH 8, 0.5 M NaCl, 1 mM TCEP and 20 mM imidazole) plus 10 mM $MgCl_2$, 10 µg/ml DNAse and 300 µg of lysozyme per gram of pellet. Cells were lysed by sonication in two cycles of 45 and 30 seconds and clarified by centrifugation at 16,000 rpm at 4°C for 15 minutes (Sorvall SS34 rotor). Cleared lysate was loaded onto a Ni-NTA affinity column prequilibrated in buffer A with protein eluted over a liner 0.05-0.5 M imidazole gradient following a 10 ml wash with 0.05 mM imidazole on an AKTA purifier (GE Healthcare). The N-terminal His_6 tag was cleaved overnight at 4°C with 3C protease during dialysis against 30 mM imidazole, 20 mM Tris-HCl pH 8, 0.5 M NaCl and 1 mM TCEP. The cleaved tag and other impurities were removed by a second round of Ni affinity chromatography. Untagged *Ax*UDT and *Ax*ABH containing fractions, as judged by absorbance at 280 nm, were pooled to a final volume of ~ 6 ml and filtered with a 0.22 µm syringe filter before loading onto a Superdex 200 16/60 gel filtration column prequillibrated in size exclusion chromatography (SEC) buffer (20 mM Tris-HCl, pH 8, 300 mM NaCl, 1mM TCEP) and eluted over one column volume.

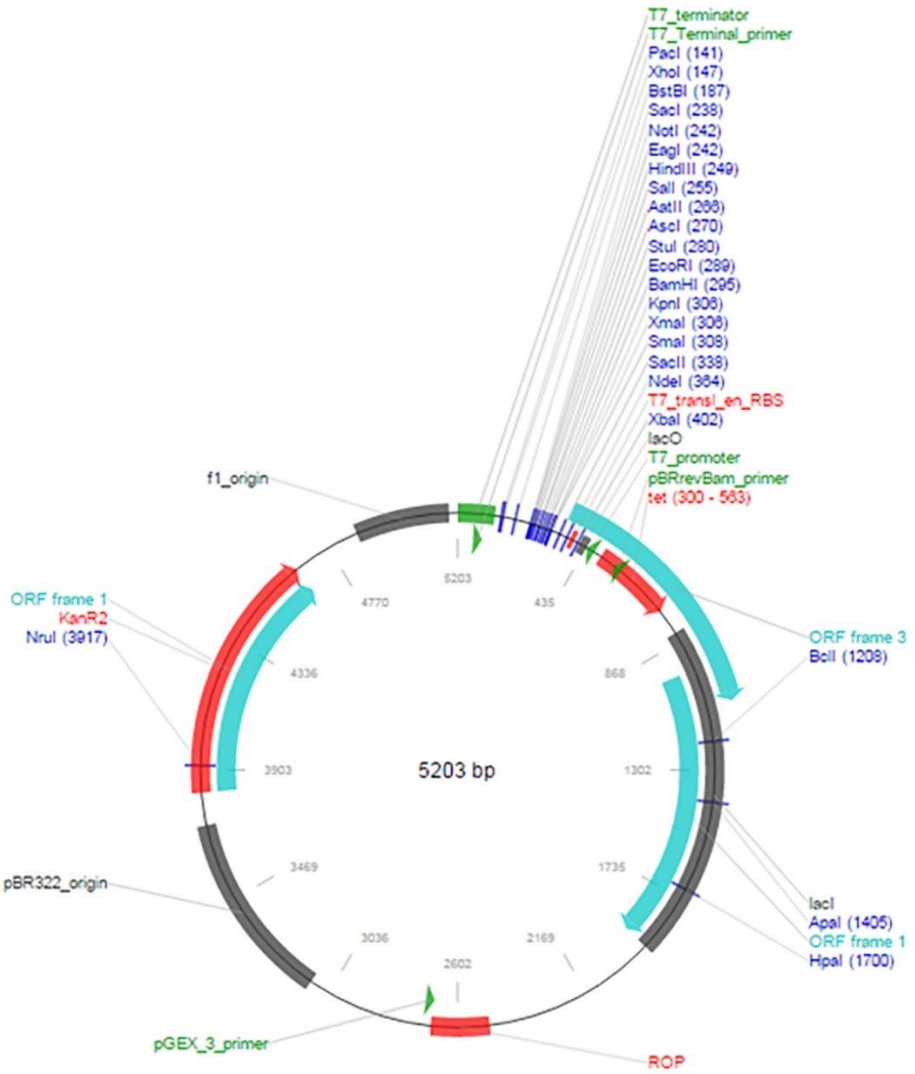


Fig. 5.2: pET47b(+) vector map.

5.2.2. Crystallisation

Purified *Ax*UDT and *Ax*ABH were concentrated to 15-20 mg/ml as quantified by absorbance at 280 nm for all crystallisation experiments. Initial crystal trials were set up using the sitting drop vapour diffusion method with a Cartesian crystal robot in 200 nL drops consisting of protein and reservoir solution from Wizard I + II, Wizard III + IV, JCSG-plus HT-96, Morpheus, PEG/ION, and Index HT commercial crystallisation screens at a ratio of 1:1. Trays were incubated at 4°C and 18°C.

5.2.2.1. UTP-glucose-1-phosphate Uridyltransferase

Initial hits for *Ax*UDT were found in Index HT A3/A5/G11 conditions (0.1 M BIS-Tris pH 5.5 and 2.0 M ammonium sulphate/ 0.1 M Hepes pH 7.5 and 2.0 M Ammonium sulphate/ 0.2 m MgCl, 0.1 M BIs-Tris pH 6.5 and 25% (w/v) PEG 3350) at 18°C, 4°C and 4°C, respectively. Index HT condition G11 was optimised using a Hamilton Star Lab liquid Handler to produce precipitant gradients (22-28% w/v PEG 3350) and pH screens (0.1 M BIS-Tris pH 6-7). Crystals were grown at 4°C in 500 nL sitting drops with protein concentrations between 5-15 mg/ml set up using a TPP mosquito LCP crystallisation robot. Long rod shaped crystals appeared after several days with the best crystals growing in 0.2 M $MgCl_2$, 24% (w/v) PEG 3350 and 0.1 M BIS-Tris pH 6.5 (Fig. 5.3).

5.2.2.2. α/β Hydrolase

Initial hits for *Ax*ABH were found in several screen conditions with 25% PEG 3350 in the presence of different divalent metals including 0.2 M magnesium chloride, 0.2 M calcium acetate and 0.2 M calcium chloride; as well as in the presence of 0.2 M lithium sulphate. A large hexagonal rod also grew in 0.1 M BIS-Tris pH 6.5, 2% Tacsimate and 20% PEG 3350 which required no further optimisation. Crystals in PEG with divalent metals were set up as described for *Ax*UDT with an increasing PEG 3350 gradient (15%-26%) in the presence of different metals (0.2 M $CaCl_2$, 0.2 M $LiSO_4$, 0.2 M $MgCl_2$ or 0.2 M NaCl) and buffers (0.1 M BIS-Tris pH 5.5-6.5). The best crystals were produced around 17% PEG 3350 with different metals giving different crystal forms. Tacsimate conditions were optimised to reproduce the long hexagonal crystals with an increasing PEG 3350 gradient (10-30%) and pH screens (0.1 M BIS-Tris pH 5.5-6.5) with 2% tacsimate. An alternative optimisation with 20% PEG 3350 and an increasing tacsimate gradient (0-25%) and a pH screen (0.1 M BIS-Tris pH5.5-6.5) was also conducted.

5.2.3. Data Collection

5.2.3.1. UTP-glucose-1-phosphate Uridyltransferase

Single crystal diffraction data from *Ax*UDT crystals were collected for molecular replacement at 100 K on beamline I24 at Diamond Light Source, UK. S-SAD data were collected at λ = 2 Å on beamline I03 from multiple *Ax*UDT crystals using inverse beam data collection over 360° with 10° wedges. Heavy atom soaks were also performed on *Ax*UDT crystals by transferring crystals into crystallisation buffer in the presence of heavy atoms (10 mM Gd, PtCn, PtCl, HgT, W or Au) overnight at 4°C. Inverse beam data were collected over 360° with 10° sweeps at the appropriate heavy atom absorption edges. Multiple datasets were collected from single crystals. Data collection parameters for all *Ax*UDT crystals are presented in table 5.1. All *Ax*UDT crystals were cryoprotected by minimising surrounding solvent around the crystal when cooling (Pellegrini et al., 2011).

Table 5.1: *Ax*UDT data collection parameters.

Crystal	*Ax*UDT MR1	*Ax*UDT MR2	*Ax*UDT S-SAD	*Ax*UDT PtCn	*Ax*UDT HgT	*Ax*UDT W	*Ax*UDT Au
Wavelength (Å)	0.9860	0.9860	2.0	1.0703	1.0077	1.2123	1.0385
Beamline	I24	I24	I03	I04	I04	I04	I04
Beam Size (µm)	10 x 10	10 x 10	50 x 20	20 x 20	20 x 20	20 x 20	20 x 20
Maximum Resolution* (Å)	1.5	1.7	3.0	2.1	2.1	2.1	2.1
Total Angular Range (°)	160	150	360	360	360	360	360
Inverse Beam Wedge (°)	N/A	N/A	10	10	10	10	10
Oscillation (°)	0.10	0.15	0.10	0.10	0.10	0.10	0.10
Exposure time per frame (s)	0.10	0.15	0.10	0.10	0.10	0.10	0.10
Transmission (%)	19	25	10	100	15	20	20

5.2.3.2. α/β Hydrolase

Single crystal diffraction data for different crystal forms of *Ax*ABH were collected at 100 K over multiple visits on beamlines I02 and I04-1 at Diamond Light Source, UK. *Ax*ABH crystals grown in tacsimate and lithium sulphate were cryo-protected by briefly soaking crystal in crystallisation buffer with 25% (v/v) ethylene glycol. *Ax*ABH crystals grown in calcium chloride, calcium acetate and magnesium chloride were cryo-protected by minimising surrounding solvent around the crystal when cooling. Data collection parameters for all *Ax*ABH crystals can be found in Table 5.2.

Table 5.2: *Ax*ABH data collection parameters.

Crystal	***Ax*ABH $CaCl_2$-1**	***Ax*ABH $CaCl_2$-2**	***Ax*ABH $LiSO_4$**	***Ax*ABH $MgCl_2$**	***Ax*ABH Tac**
Wavelength (Å)	0.9795	0.9200	0.9800	0.9795	0.9200
Beamline	I02	I04-1	I02	I02	I04-1
Beam Size (µm)	82 x 21	60 x 50	20 x 20	82 x 21	60 x 50
Maximum Resolution* (Å)	1.6	1.7	1.5	2.0	1.6
Total Angular Range (°)	160	250	140	170	120
Oscillation (°)	0.10	0.10	0.10	0.10	0.10
Exposure time per frame (s)	0.10	0.10	0.10	0.10	0.10
Transmission (%)	30	11	30	10	17

5.2.4. Structure Solution and Refinement

5.2.4.1. UTP-glucose-1-phosphate Uridyltransferase

Molecular replacement was performed using Phaser (McCoy et al., 2007) and Molrep (Vagin and Teplyakov, 2010a) with various sequence and secondary structure homologues and using the automated molecular replacement pipeline MrBUMP (Keegan and Winn, 2007, Keegan and Winn, 2008). S-SAD and heavy atom soaked data were processed using XDS (Kabsch, 2010b) with

data scaled together using XSCALE and merged with AIMLESS (Evans and Murshudov, 2013) to assess the presence of anomalous signal.

5.2.4.2. α/β Hydrolase

Diffraction data for all crystal forms were processed with the Xia2 3daii pipeline at Diamond light source (Kabsch, 2010b, Winter et al., 2013). All structures were solved by molecular replacement with Molrep (Vagin and Teplyakov, 2010a). *Ax*ABH-Li was solved using chain A of a putative lipase from *Acinetobacter brumannii* (PDB code: 4OPM, swquence identity 25%) and the model built with Buccaneer (Cowtan, 2012) using the Molrep molecular replacement model as the PDB input. Further refinement of the structures was performed with REFMAC5 (Murshudov et al., 2011), with model building carried out in COOT (Emsley et al., 2010). Model validation was performed using Molprobity (Chen et al., 2010). *Ax*ABH-Ca1, *Ax*ABH-Ca2, *Ax*ABH-Mg and *Ax*ABH-Tac were solved using the refined *Ax*ABH-Li structure as the search model. Figures of all structures were prepared with PyMol version 1.3r1 (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC).

5.3. Results

5.3.1. Expression and Purification

Expression trials of both AxUDT and AxABH showed high levels of expression in BL21 *E. coli* at molecular weights around 34 KDa and 39 KDa, respectively (Fig. 5.3). Purification of both proteins was relatively straightforward.

5.3.1.1. UTP-glucose-1-phosphate Uridyltransferase

His tagged AxUDT was eluted from the Ni NTA column at high purity (Fig. 5.4) with the majority of impurities removed in the initial wash to produce a single well defined peak at 280 nm. Final impurities were removed to give near 100% purity after cleavage of the His tag and gel

filtration (Fig. 5.5) to give approximately 8 mg of protein as judged by absorbance at 280 nm. Gel filtration showed a single symmetrical peak suggesting *Ax*UDT exists in solution as a monomer (Fig. 5.6).

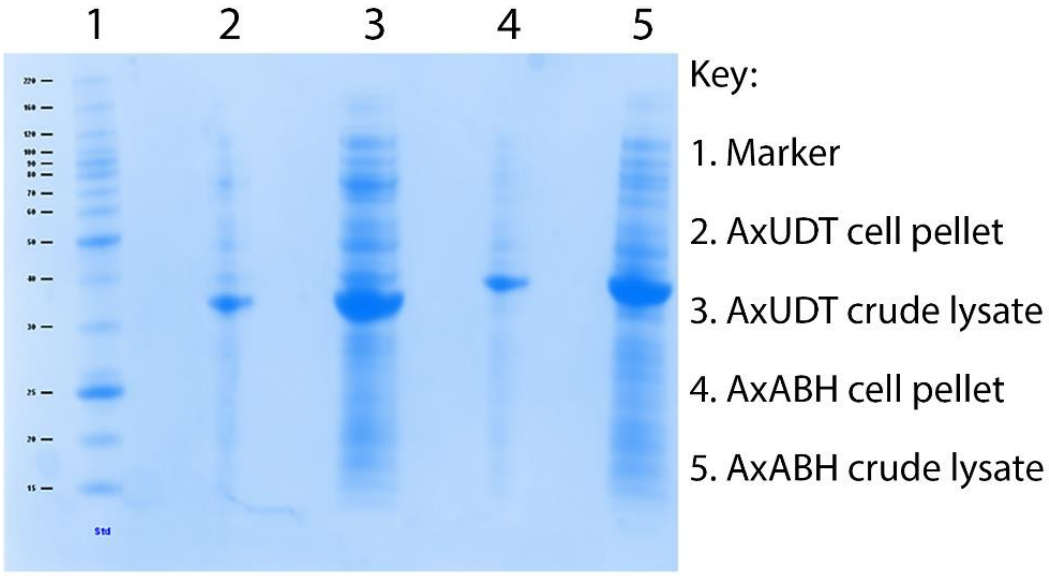


Fig. 5.3: AxUDT and AxABH expression trial SDS PAGE

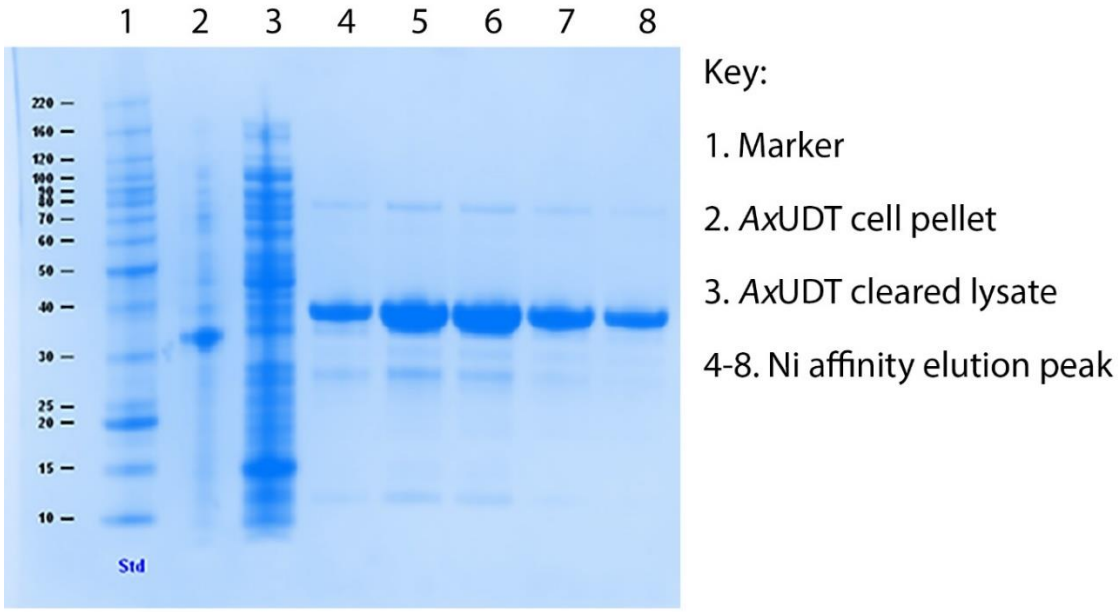


Fig. 5.4: *Ax*UDT Nickle affinity elution SDS PAGE.

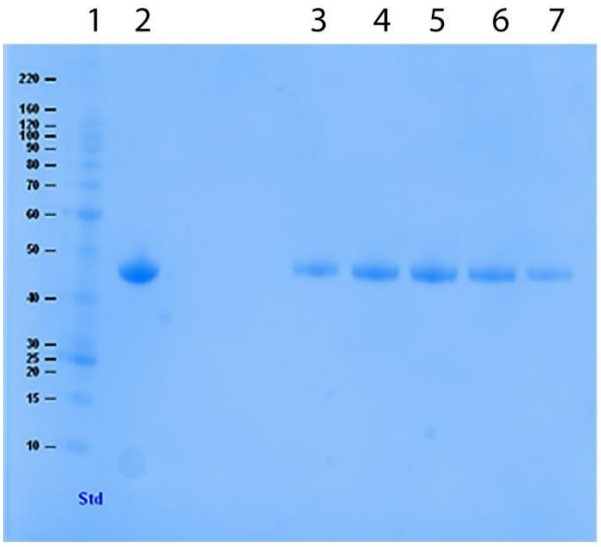


Fig. 5.5: *Ax*UDT gel filtration SDS PAGE

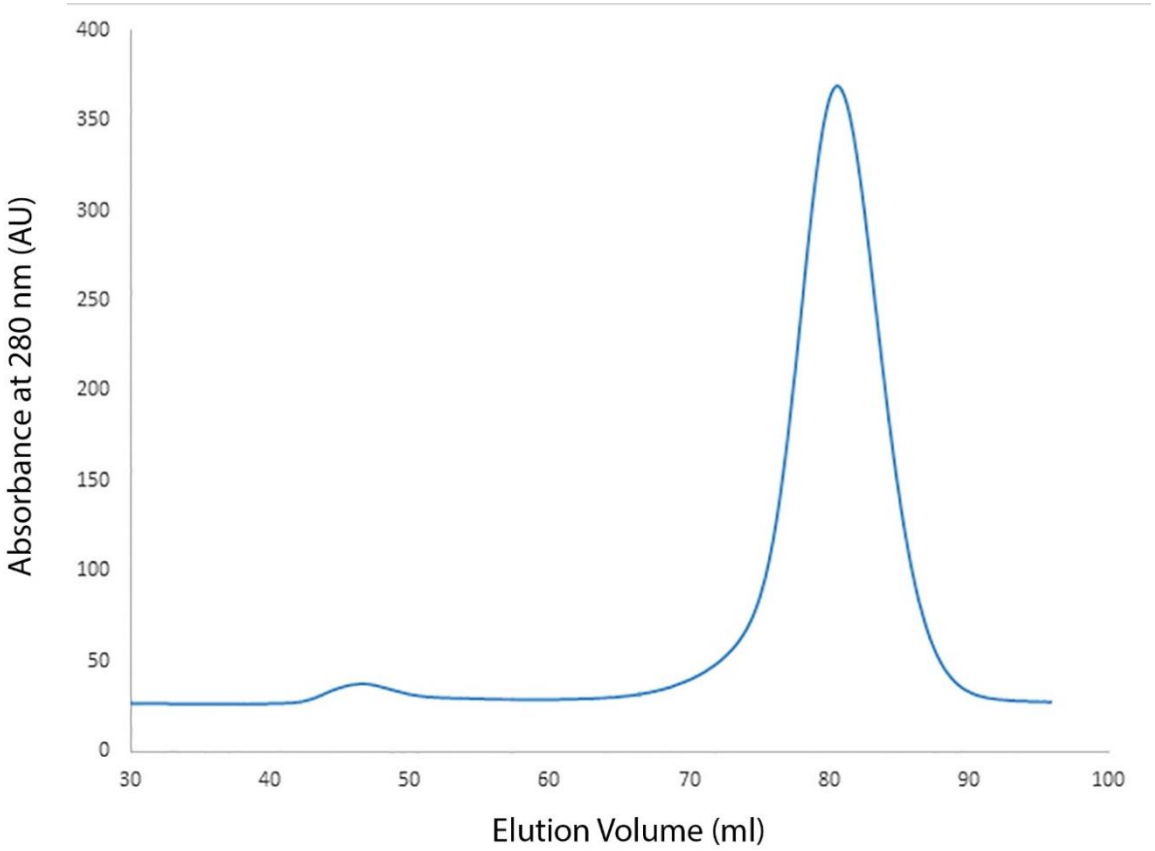


Fig. 5.6: *Ax*UDT gel filtration elution peak

5.3.1.2. α/β Hydrolase

Initial Ni affinity purification of *Ax*ABH showed some impurities but an overwhelming majority of the protein present was *Ax*ABH (Fig. 5.7, lanes 4-6). Most impurities were removed following the second round of Ni affinity purification (Fig. 5.7, lane 9) and near 100% pure *Ax*ABH after gel filtration (Fig. 5.7, lane 10) to give 25 mg of protein as judged by absorbance at 280 nm. A single broad symmetrical peak from gel filtration suggests *Ax*ABH exists as a monomer in solution (Fig. 5.8).

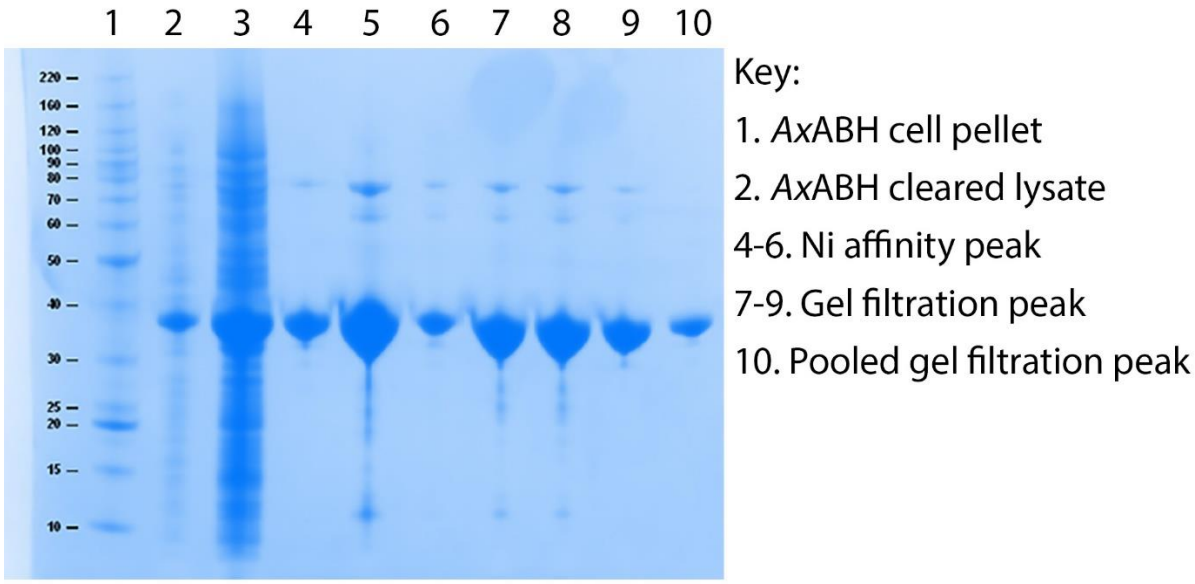


Fig. 5.7: *Ax*ABH nickle ffinity and gel filtration SDS PAGE

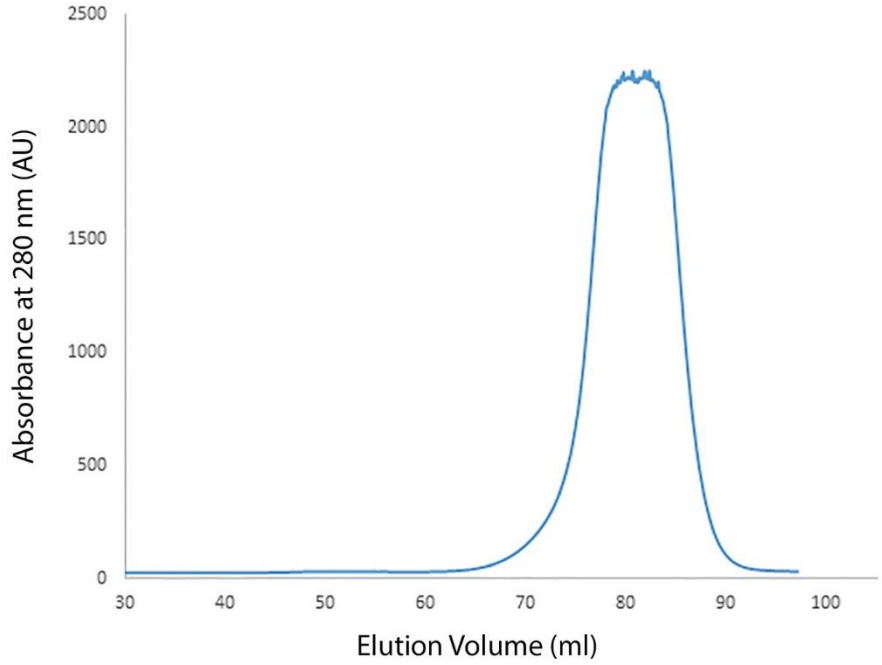


Fig. 5.8: *Ax*ABH gel filtration elution peak

5.3.2. Crystallisation

5.3.2.1. UTP-glucose-1-phosphate Uridyltransferase

Initial crystal hits for *Ax*UDT grew primarily as thin rod shaped crystals with some frayed tree shaped crystals as shown in Figure 5.9. Other showers of crystals were observed but not optimised. Ammonium sulphate and PEG 3350 conditions (Fig. 5.9a-b and 5.9c, respectively) outlined earlier were optimised, with the PEG 3350 condition giving reproducible long rod shaped crystals around 24% PEG 3350 (Fig. 5.9c).

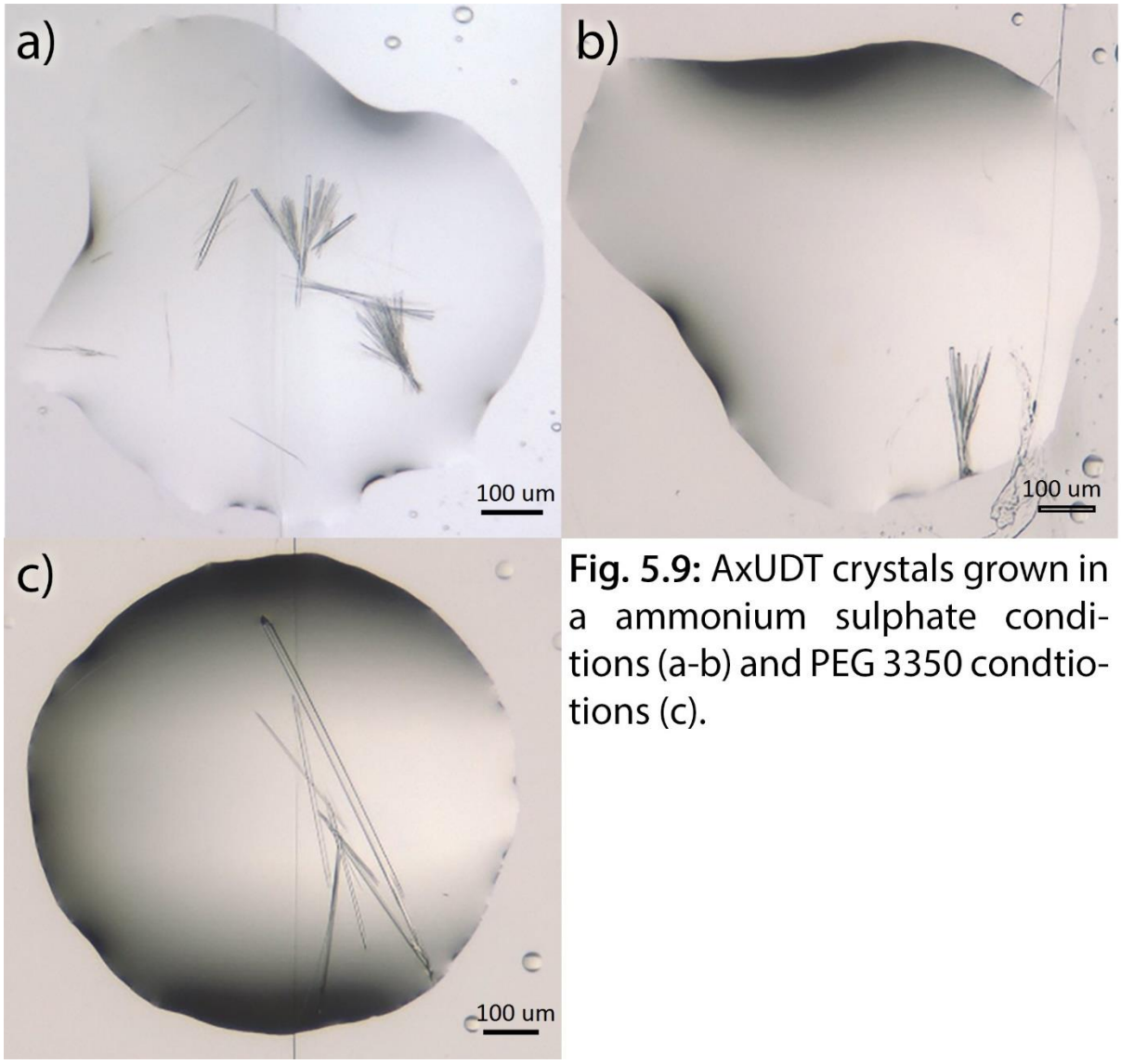


Fig. 5.9: AxUDT crystals grown in a ammonium sulphate conditions (a-b) and PEG 3350 condtiotions (c).

5.3.2.2. α/β Hydrolase

Several crystal forms were produced in optimised PEG 3350 conditions with varying divalent metals and lithium sulphate (Fig. 5.10a-c) which proved relatively reproducible. The long hexagonal crystal from initial screening (Fig. 5.10d) proved more difficult to reproduce with scaled up drop sizes failing to reproduce the same crystal form. Other crystals similar to those observed in the PEG 3350/divalent/lithium conditions were observed but never another large hexagonal rod.

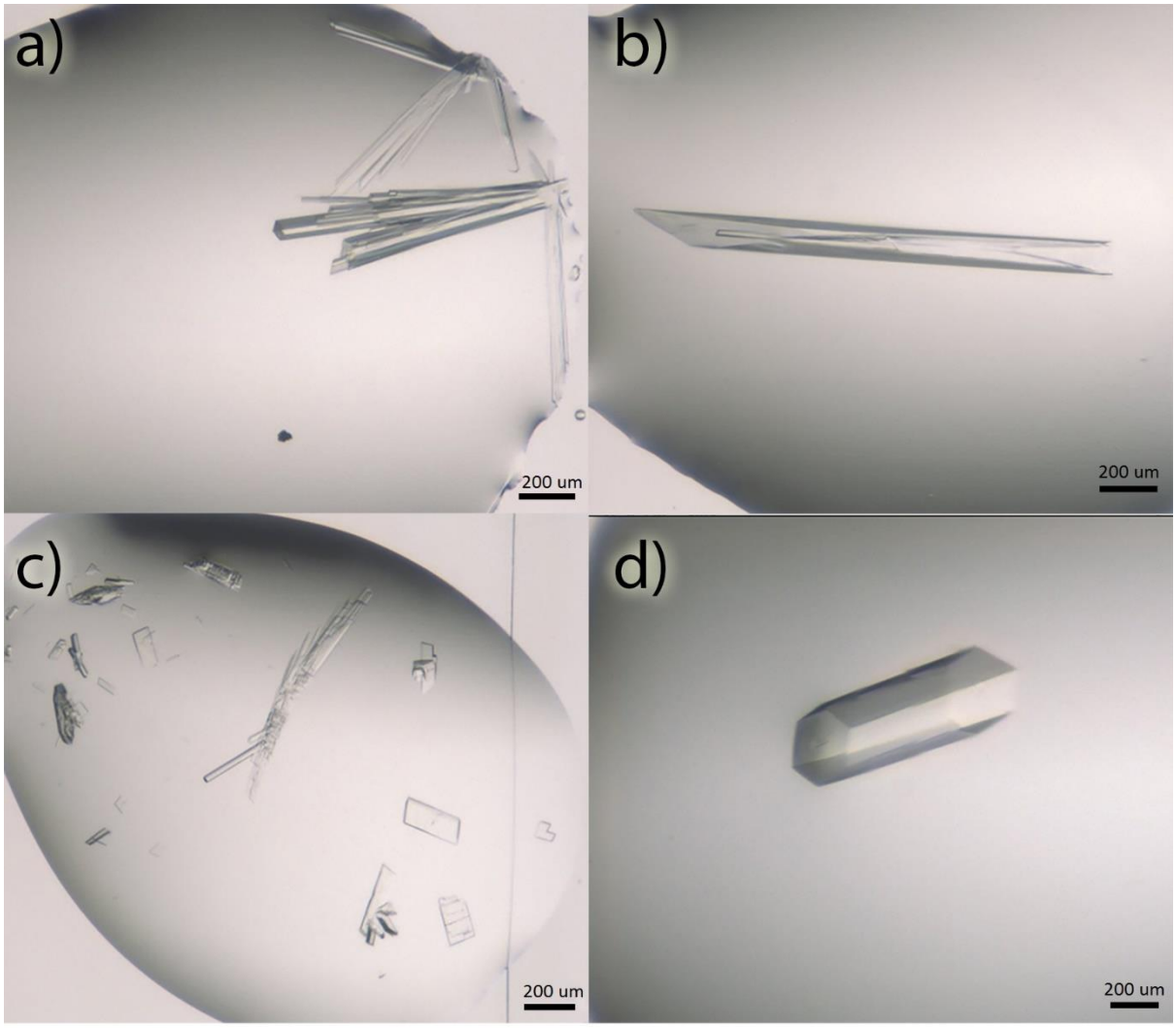


Fig. 5.10: AxABH crystals grown in PEG 3350 conditions with lithium sulphate (a), calcium acetate (b), calcium chloride (c) and tacsimate (d).

5.3.3. Structure Solution and Refinement

5.3.3.1 UTP-glucose-1-phosphate Uridyltransferase

Despite a high level of sequence identity (49%: 3JUJ) *Ax*UDT could not be phased by molecular replacement with all homology models tried using Molrep or Phaser. Models with high secondary structure homology from HHpred (Soding et al., 2005) also failed to produce a sensible rotation and translation function. Models were trimmed manually to remove flexible loops and leave the conserved core based on alignment of multiple homology models, but still failed to produce any

contrast in the molecular replacement solutions. Molecular replacement models generated with CHAINSAW (Stein, 2008) and MrBUMP (Keegan and Winn, 2008) also failed to produce a sensible solution. Analysis of the solvent content shows *Ax*UDT-MR1 and *Ax*UDT-MR2 to have between 8-10 molecules in the asymmetric unit which may explain the difficulty of finding a sensible solution by molecular replacement. Processing determined the spacegroup as $P2_1$ with a unit cell of a = 80.2, b = 89.5 and c = 179.0 with α = 90 β = 100.1 and γ = 90. Closer inspection of *Ax*UDT-MR1 also showed the presence of multiple crystal lattices which could account for further difficulty in finding a solution.

S-SAD data were collected on nine *Ax*UDT crystals at 2 Å wavelength. Of the nine crystals two showed multiple lattices in the diffraction pattern, two suffered severe radiation damage within a few hundred images (multiplicity ~ 0.5 and completeness ~ 21%) and one showed little to no diffraction. Shorter exposure times and reduced transmission should be considered to reduce radiation damage if the diffraction experiment is repeated. Multiple datasets were collected from the four remaining crystals but no discernible anomalous signal was observed in individual datasets, with an anomalous slope between 0.85-1.10 and anomalous completeness between 50-85%. The low anomalous completeness is likely to be due to the low symmetry space group (*C*2/*P*2) and the low anomalous multiplicity (~1.4). Scaling multiple datasets from the same crystal in all possible combinations with XSCALE to increase anomalous completeness showed no improvement in the anomalous slope and SHELX failed to identify the anomalous substructure. Combining the data from multiple crystals was not attempted as indexing solutions for different crystals gave significantly different solutions (Table 5.A1). Although *Ax*UDT crystals were not in an optimum space group (*P*2, *C*2 or *P*1) to allow high multiplicity anomalous data to be collected the inability to generate significant anomalous signal when combining several datasets demonstrated the limitations of S-SAD phasing on standard MX beamlines (Doutch et al., 2012) and the need for a specialised beamline such as I23.

Heavy atom soaking also failed to produce a solution. Of the soaks attempted PtCn, HgT, W and Au soaked crystals produced diffraction, however, all but W soaked crystals showed weak diffraction to ≤ 4 Å and failed to process. Inverse beam datasets for each heavy atom soak were scaled with XSCALE (Kabsch, 2010b) but showed no detectable anomalous signal. Multiple datasets from the same crystal were scaled together using XSCALE in all possible combinations but data proved non-isomorphous and detrimental to measuring any anomalous signal. Data were reprocessed with XDS (Kabsch, 2010b) with fewer frames to reduce the effects of radiation damage but showed no improvement in anomalous signal.

5.3.3.2. α/β Hydrolase

High to medium resolution structures of *Ax*ABH in the presence of lithium, magnesium, tacsimate and two crystals forms with calcium were obtained at 1.81 Å, 2.0 Å, 1.76 Å, 2.12 Å and 1.72 Å resolution respectively. Data processing, refinement and validation statistics for all structures are presented in Table 5.3. *Ax*ABH shows the typical α/β hydrolase fold with a twisted eight stranded β-sheet flanked by several α-helices (Fig. 5.11b, coloured blue). After β-sheet six *Ax*ABH has a five helix domain (α4-8) constituting an additional domain of unknown function (DUF). The DUF consist of a parallel helix turn helix motif made up of α4 and α5 (HTH1) followed by a second longer parallel helix turn helix motif from α6 and α7 (HTH2) and finally a single long α-helix (α8) spanning 28 residues which passes behind HTH1 (Fig. 5.11b, coloured orange). This DUF forms a hydrophobic pocket, presumably the active site, which brings the substrate into the vicinity of the catalytic Nucleophile-His-Acid triad formed by the conserved α/β hydrolase fold (Marchot and Chatonnet, 2012). The active site was determined based on the presence of a bound ligand which appears to be 2,2-Bis(hydroxymethyl)-2,2',2"- nitrilotriethanol (BIS-tris) from the crystallisation condition based on the shape of the difference density (Fig. 5.12a-b). LigPlot+ generated an active site consisting of Tyr59, Asp127, Tyr128, Gly162, Pro165, Phe184, Leu229, Tyr232, Phe263 and Phe264 (Fig. 5.12c). However, *Ax*ABH doesn't conform to the conserved catalytic triad with a phenylalanine (Phe264) in

the acid position and an aspartate (Asp127) in the nucleophilic position (Fig. 5.12a-b, coloured cyan and green, respectively). The highly conserved histidine residue is present (Fig. 5.12a-b, coloured red). This suggests *Ax*ABH may be an inactive hydrolase.

Table 5.3: Crystallization, crystallographic data collection and structure refinement statistics for *Ax*ABH. Values in parentheses refer to the highest resolution bin.

Structure	***Ax*ABH-Li**	***Ax*ABH-Ca1**	***Ax*ABH-Ca2**	***Ax*ABH-Mg**	***Ax*ABH-Tac**
Data Collection					
Space Group	*C*2	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Cell Axes (Å)	99.06 x 78.42 x 46.27	46.13 x 81.74 x 157.50	91.09 x 83.22 x 134.20	58.43 x 168.49 x 69.62	81.13 x 94.75 x 170.00
Angles (°)	90.00 x 106.64 x 90.00	90.00 x 96.11 x 90.00	90.00 x 106.94 x 90.00	90.00 x 102.19 x 90.00	90 x 90 x 90
Wavelength (Å)	0.9795	0.9795	0.9200	0.9795	0.9200
Resolution (Å)	39.33-1.81	56.54-2.12	60.83-1.72	68.05-2.00	58.69-1.76
Unique Reflections (#)	28570 (2091)	60340 (3691)	194851 (14750)	87387 (6402)	129182 (9384)
Redundancy	2.7 (2.6)	2.5 (2.8)	4.8 (4.7)	3.4 (3.3)	3.9 (4.5)
R_{pim} (%)	5.9 (47.8)	4.3 (39.5)	4.2 (34.6)	5.8 (32.4)	3.6 (38.7)
I/σ(I)	10.2 (1.9)	13.1 (2.1)	13.6 (2.3)	8.7 (2.4)	16.5 (2.0)
Completeness (%)	97.3 (96.0)	91.1 (75.4)	98.7 (95.9)	98.4 (98.8)	98.4 (99.3)
CC1/2	0.995 (0.691)	0.997 (0.768)	0.997 (0.730)	0.998 (0.853)	0.999 (0.734)
Refinement					
Molecules/AU	1	4	6	4	4
R_{work}/R_{free} (%)	19.7/24.0	21.6/27.4	22.2/25.3	20.2/25.0	20.6/24.7
Rmsd Bond Length (Å)	0.017	0.015	0.020	0.015	0.018
Rmsd Bond Angles (°)	1.772	1.719	1.963	1.712	1.833
Average B factor ($Å^2$)	30.8	41.2	19.4	34.4	29.4
Solvent Content (%)	51.5	43.4	48.5	50.1	48.8
Ramachandran Statistics (%)					
Favoured Regions	97.26	95.49	98.20	96.35	96.90
Allowed Regions	2.40	4.42	0.74	3.65	2.93

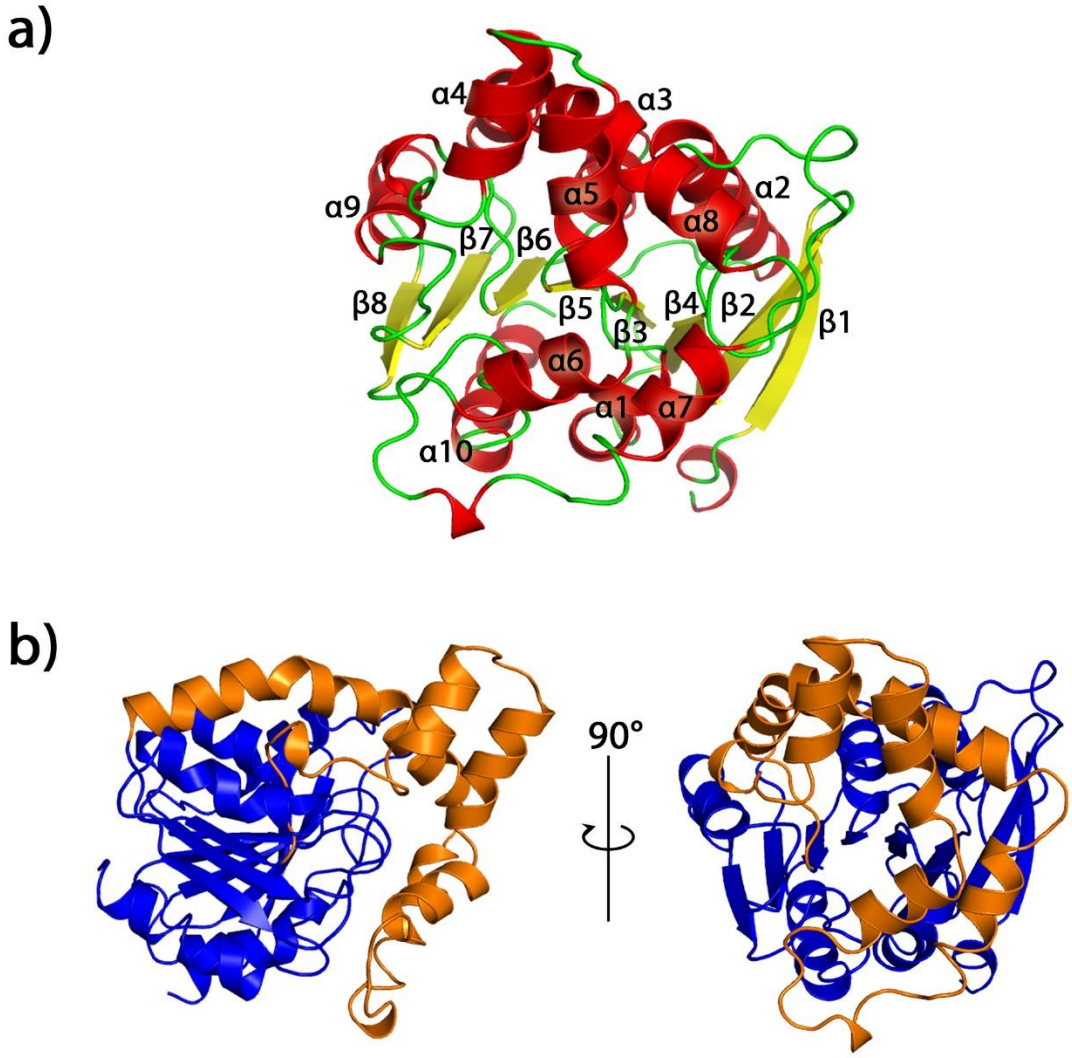


Fig. 5.11: The Structure of *Ax*ABH in cartoon form (a) with α-helices coloured in red, β-sheets coloured in yellow and loops coloured in green. (b) AxABH coloured to show the conserved α/β fold in blue and the novel domain of unknown function (DUF) in orange.

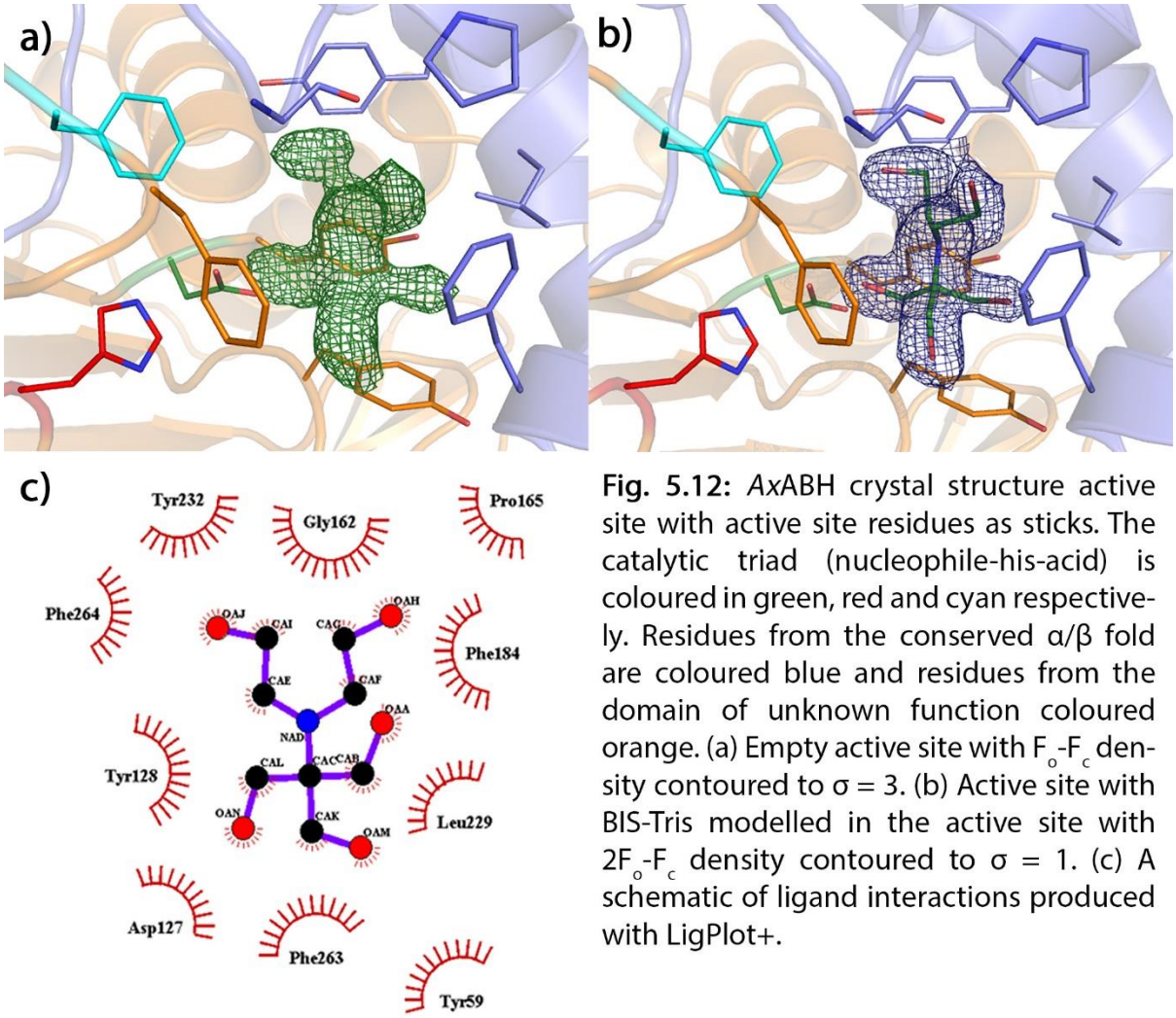


Fig. 5.12: *Ax*ABH crystal structure active site with active site residues as sticks. The catalytic triad (nucleophile-his-acid) is coloured in green, red and cyan respectively. Residues from the conserved α/β fold are coloured blue and residues from the domain of unknown function coloured orange. (a) Empty active site with F_o-F_c density contoured to σ = 3. (b) Active site with BIS-Tris modelled in the active site with $2F_o$-F_c density contoured to σ = 1. (c) A schematic of ligand interactions produced with LigPlot+.

5.4. Discussion

5.4.1. *Ax*ABH Crystal Forms

Crystals from optimisation of *Ax*ABH with different divalent metals produced four different crystal forms, (1) space group C2 with one molecule in the asymmetric unit in the presence of lithium (Fig. 5.13a); (2) space group $P2_1$ with 4 molecules in the asymmetric unit in the presence of calcium (Fig. 5.13b); (3) space group $P2_1$ with 4 molecules in the asymmetric unit in the presence of magnesium (Fig. 5.13c) and (4) space group $P2_1$ with 6 molecules in the asymmetric unit in the presence of calcium (Fig. 5.13e). A fifth crystal form was produced from initial screening in the presence of tacsimate in space group $P2_12_12_1$ with 4 molecules in the asymmetric unit (Fig. 5.13d). Of these

crystal forms the tacsimate condition would likely have been the most effective crystal for sulphur phasing on I23 with 90 degrees of data needed for a complete dataset (compared with 180 degrees for all other crystal forms) and a Bijvoet ratio of 0.32%, 1.37%, 2.58% and 4.06% at 1 Å, 2 Å, 3 Å and 4 Å, respectively.

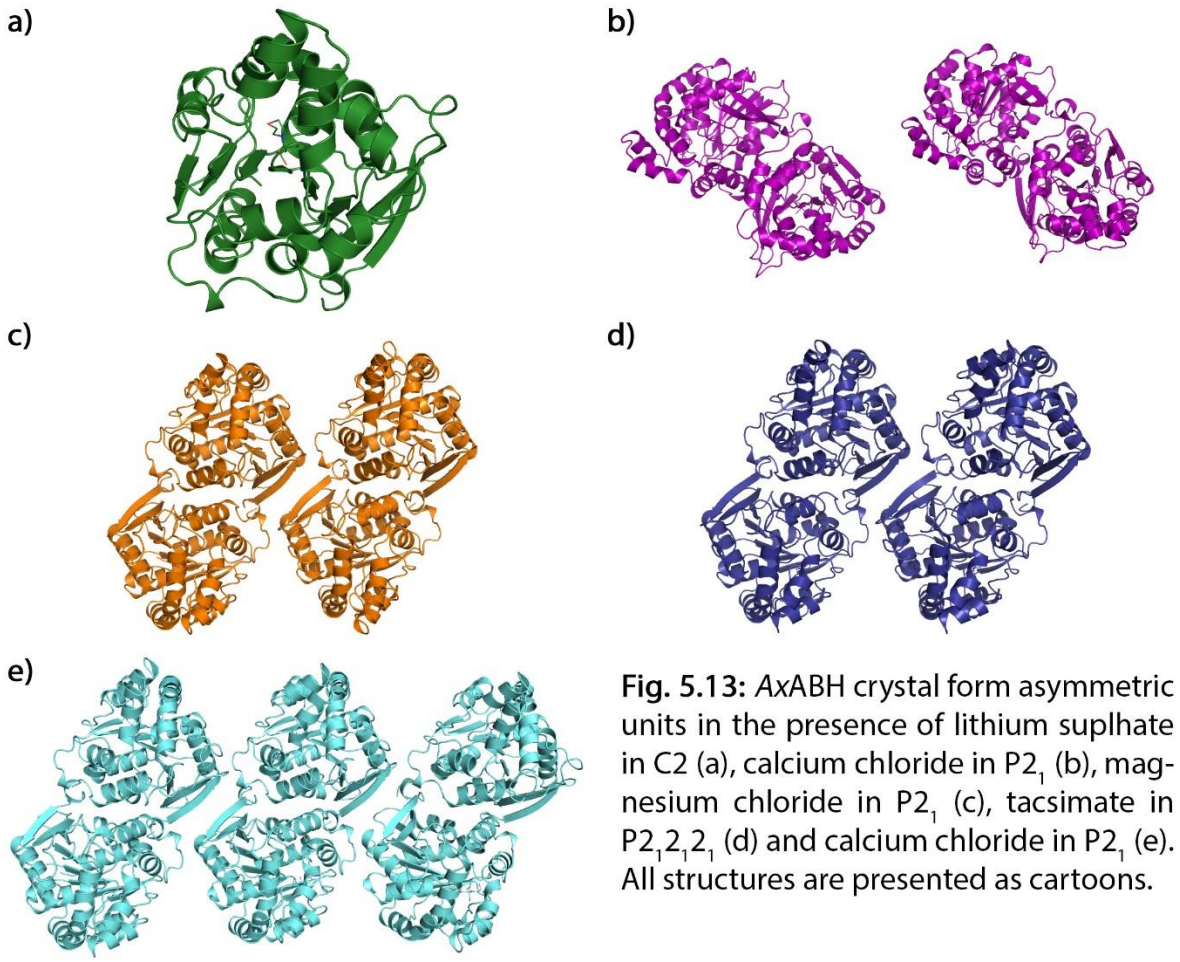


Fig. 5.13: *Ax*ABH crystal form asymmetric units in the presence of lithium suplhate in C2 (a), calcium chloride in $P2_1$ (b), magnesium chloride in $P2_1$ (c), tacsimate in $P2_12_12_1$ (d) and calcium chloride in $P2_1$ (e). All structures are presented as cartoons.

Superposition of the A chains of all crystal forms with the single molecule *Ax*ABH-Li structure showed little change in the overall fold of the protein with both the conserved α/β fold and DUF showing little deviation. The only observable conformational change is in HTH1 where *Ax*ABH-Li (green) and *Ax*ABH-Ca2 (cyan) structures show a displacement of approximately 3.6 Å from *Ax*ABH-Ca1 (purple), *Ax*ABH-Tac (blue) and *Ax*ABH-Mg (orange) (Fig. 5.14). Closer inspection of the active site shows Gly162 in this region has moved in relation to the position of the BIS-Tris ligand in the active site. AxABH-Li and AxABH-Ca2's A chain show a vertical orientation of BIS-Tris relative to

Asp127 (Fig. 5.15, coloured green and cyan, respectively) while AxABH-Ca1, AxABH-Tac and AxABH-Mg have a horizontal orientation (Fig. 5.15 coloured purple, blue and orange, respectively). The orientation of BIS-Tris in the active site also varies between chains within the same structures so may be the result of crystal packing or BIS-tris being a poor fit for the active site. Ligand $2F_o$-F_c density at σ = 1 for each crystal form is presented in Figure 5.16.

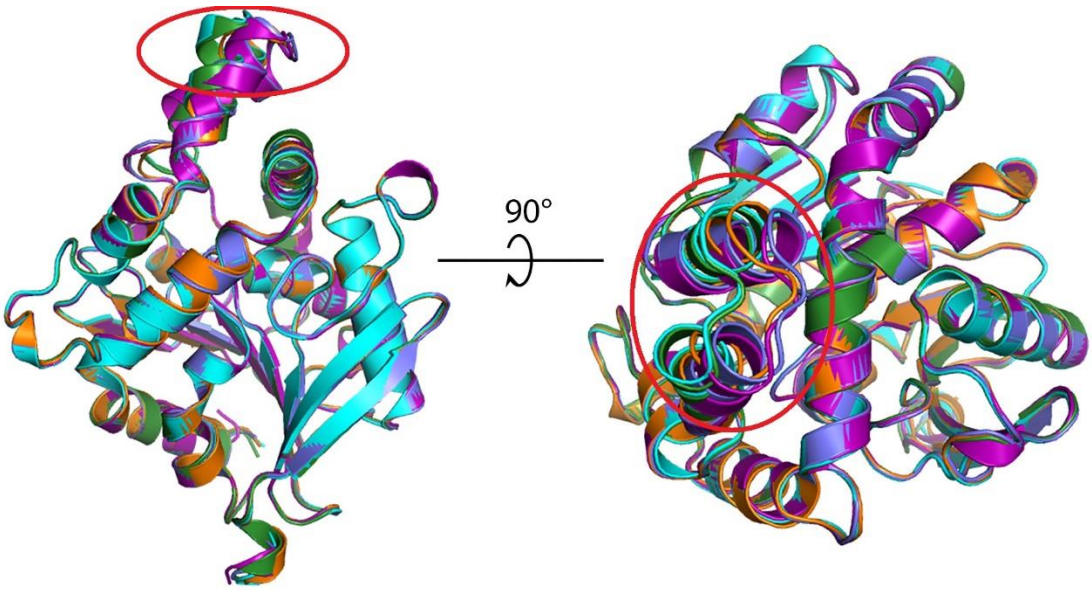


Fig. 5.15: Superposition of *Ax*ABH structure A chains from lithium sulphate (green), Calcium chloride 1 (purple), magnesium chloride (orange), tacsimate (blue) and calcium chloride 2 (cyan).

5.4.2. *Ax*ABH Structure and Function

Sequence alignment with BLASTp (Altschul et al., 1990) generated several high percentage homologues, up to 87% sequence homolgy at best, all of which were esterases/lipases which act on carboxylic esters and have a conserved ser-his-glu/asp as their nucleophile-his-acid catalytic triad. However, *Ax*ABH lacks both nucleophile and acid residues and shows poor structural alignment with carboxylaesterase structures deposited in the PDB.

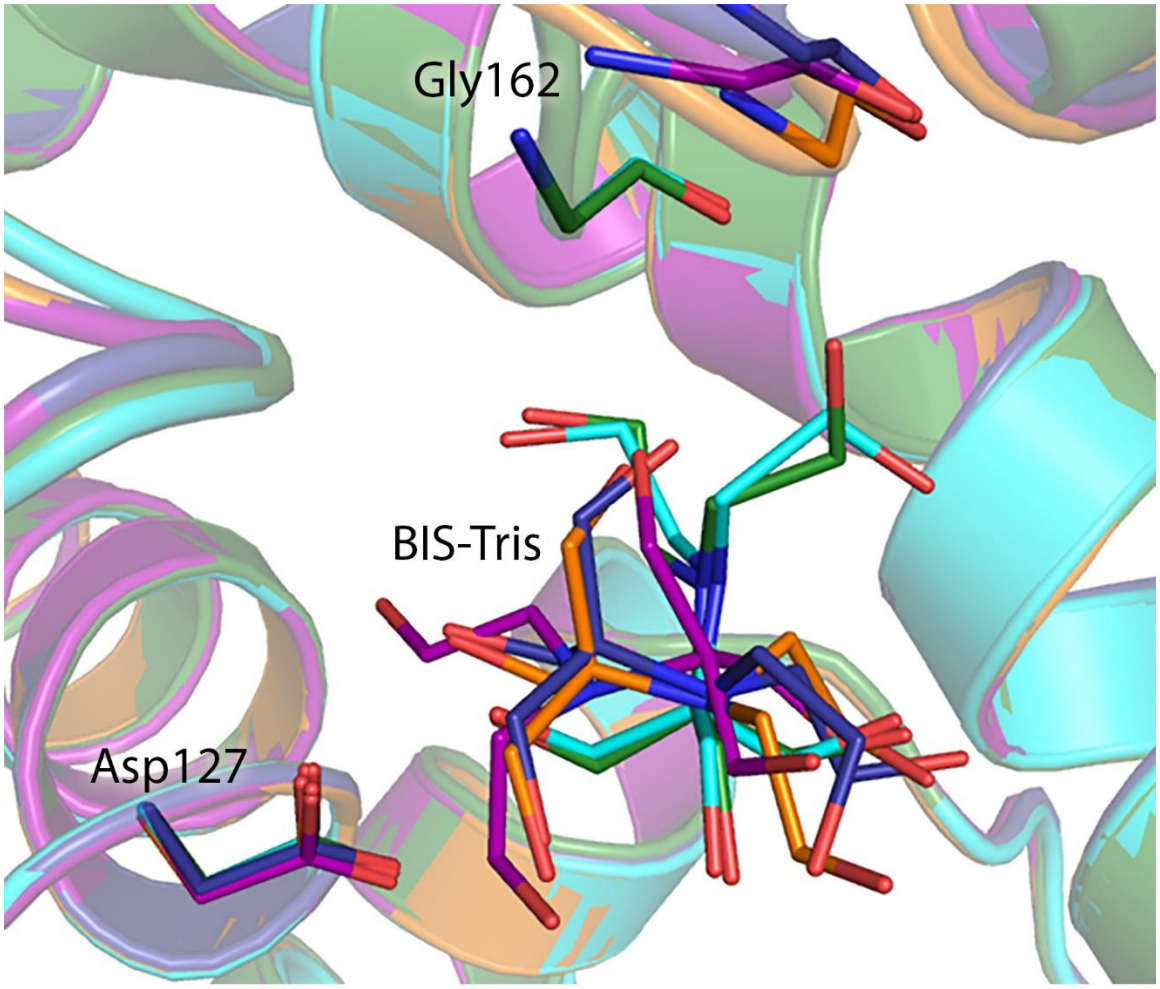


Fig. 5.15: AxABH active site superposition with BIS-Tris and residues as sticks. Structures, ligands and residues of AxABH in the presence of lithium sulphate, calcium chloride (1), magnesium chloride, tacsimate and calcium chloride (2) are coloured green, purple, orange, blue and cyan, respectively.

Secondary structure alignment with PDBe (Krissinel and Henrick, 2004) produced homologous models with secondary structure similarity up to 78% and an RMSD of 1.74. Structures included a CIB like hydrolase (PDB code: 3BDI), C-C hydrolase (PDB code: 4LYD and 4LXH), serine hydrolase (PDB code: 1J1I), carboxylesterase (PDB code: 4CCY) and haloalkane dehydrogenase (PDB code: 3A2N) but alignment corresponded to the conserved α/β hydrolase fold with none of the identified structures showing a similar fold to *Ax*ABH's DUF. Comparison of Protein Active Site Structures Server (Powers et al., 2011) was used to identify similar functional epitopes and active sites based on the residues binding the ligand. The strongest match was a phenylethylamine oxidase (PDB code: 2CFW) with a 36.6% similarity but alignment of the structures showed little agreement

between the protein folds. Analysis of the crystal formed showed a common dimer form between some crystal forms (Fig. 5.13c-e). The Proteins, Interfaces, structures and Assemblies (PISA) server (Krissinel and Henrick, 2007) identified a stable dimeric assembly with a dissociation constant of 4.5 kcal/mol in *Ax*ABH between chains E and F of the Mg-ABH structure. However, analysis of the C and D chains showed a dissociation constant of -1.2 kcal/mol. Overall PISA found no strong indication that *Ax*ABH forms a dimer in solution, in agreement with the monomeric state observed in gel filtration purification (Fig. 5.8).

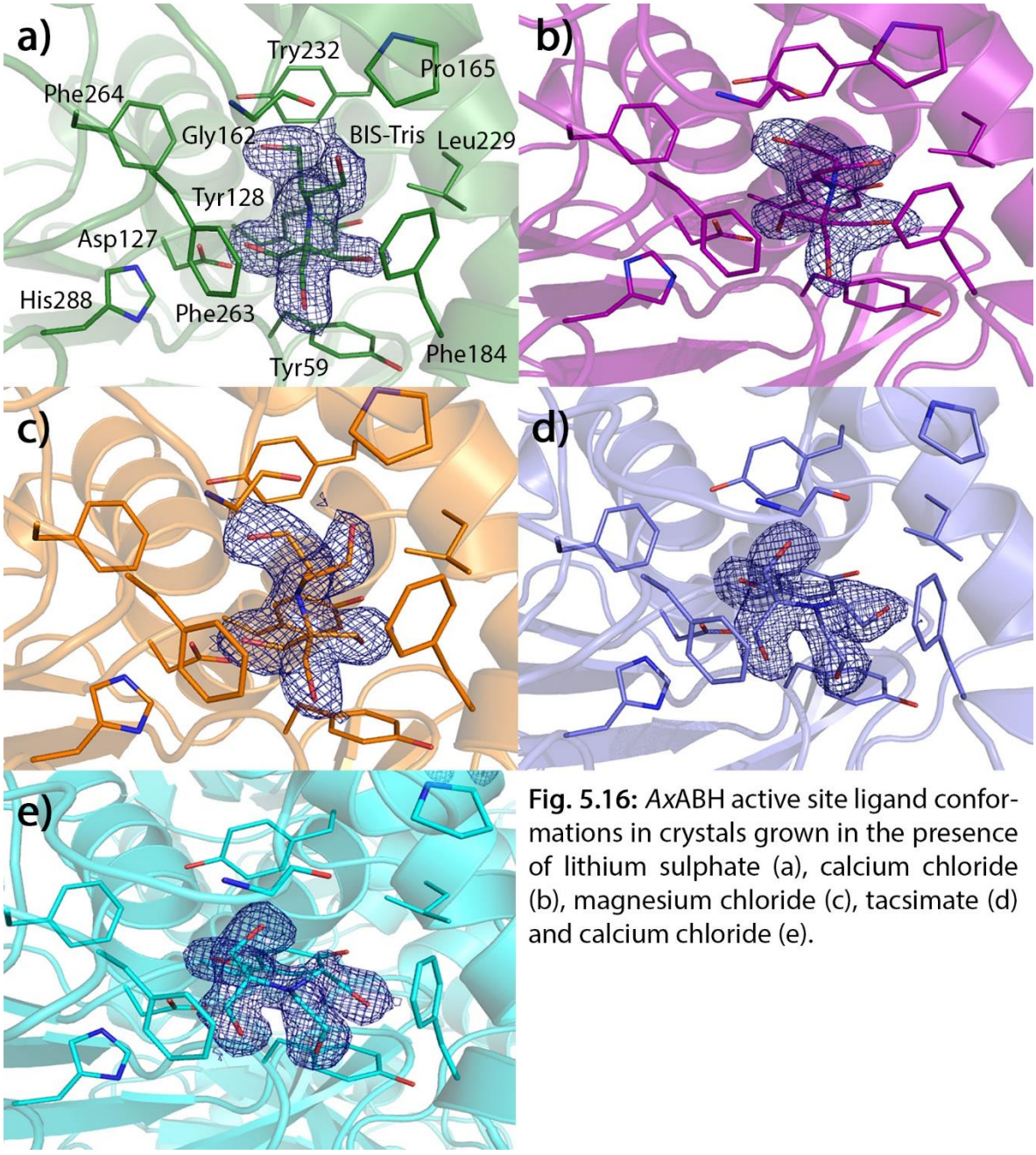


Fig. 5.16: *Ax*ABH active site ligand conformations in crystals grown in the presence of lithium sulphate (a), calcium chloride (b), magnesium chloride (c), tacsimate (d) and calcium chloride (e).

5.5. Conclusions

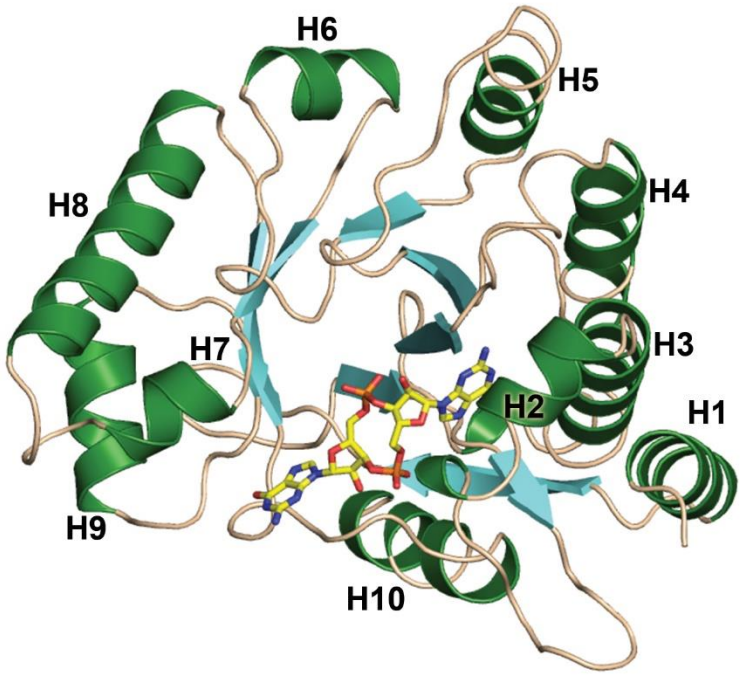
Diffracting protein crystals were successfully produced for both *Ax*UDT and *Ax*ABH for long wavelength phasing experiments on I23 but could not be tested due to technical difficulties with the beamline. *Ax*UDT makes a good case for S-SAD on I23 with the failure to produce a solution by molecular replacement due to large numbers of molecules in the asymmetric unit and non-isomorphism making combinations of datasets for SAD phasing impossible. With between 80-100 sulphur atoms in the asymmetric unit *Ax*UDT should produce a very strong anomalous signal. *Ax*ABH was successfully solved by molecular replacement but could still prove a valuable test subject for the limits of S-SAD with its low sulphur content. The function of *Ax*ABH remains unknown with the novel DUF domain showing little similarity to sequence, secondary structure and liganding residue homologues.

5.6. Appendix

Table 5.A1: *Ax*UDT crystal forms from S-SAD data collection.

Crystal form	Space Group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)
1	C2	80.5	90.3	137.3	90.0	105.8	90.0
2	P1	63.0	63.8	120.5	89.1	83.8	77.3
3	P2	80.5	90.9	179.9	90.0	99.9	90.0

Chapter 6 – Structural Insight into the Mechanism of EAL Domains



6.1. Introduction

6.1.1 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a gram negative rod bacteria with a polar flagella from the family *Pseudomonadaceae.* It is oxidase positive, based on the presence of cytochrome c oxidase (Gaby and Hadley, 1957), and largely aerobic and non-fermentative; but has been found to be able to ferment nitrate and arginine (Vander Wauven et al., 1984, Palmer et al., 2007). *P. aeruginosa* presents in various colours depending on secretion of different pigments including blue-green, yellow- green and red-brown from pyocycanin, pyoverdine and pyorubin, respectively. *P. aeruginosa* strains are widespread throughout the world (Römling et al., 2005) and can inhabit soil, water, plants and animals (including humans) (Tolker-Nielsen, 2014).

6.1.1.1. *Pseudomonas aeruginosa* Infection

P. aeruginosa is an opportunistic pathogen, rarely causing disease without prior damage to a protective layer such as the skin or mucus membrane. Infection is also prevalent in immunodeficient individuals such as HIV/AIDS patients, immunosuppressed patients, post-transplant or following chemotherapy and in cystic fibrosis sufferers. According to the Centre for Disease Control (CDC) an estimated 51,000 healthcare-associated *P. aeruginosa* infections occur per year in the USA with over 13% being multidrug-resistant strains and approximately 400 deaths per year resulting from infections (CDC, 2014). Owing to *P. aeruginosa's* prevalence in the natural world prevention presents a considerable challenge to hospital care. All *P. aeruginosa* infections are considered treatable and potentially curable, however, acute fulminant infections such as bacteremic pneumonia, sepsis and meningitis have extremely high mortality rates.

Respiratory tract infection by *P. aeruginosa* is one of the most common infections in cystic fibrosis sufferers, although the latest data suggest a decrease in the prevalence in patients from a high of ~ 60% of sufferers in 1988 to ~ 50% in 2013 (Fig. 6.1) (Cystic Fibrosis Foundation Patient

Registry. Annual Data Report 2013. Cystic Fibrosis foundation, Bethesda, MD). Studies suggest that this decrease is the result of early intervention with antibiotics to reduced chronic colonisation (Davidson et al., 2012) and a general improvement in understanding the clinical impact, diagonosis and treatment of cyctic fibrosis sufferers with *P. aeruginosa* infections. Infection is thought to occur through natural contact with lakes, streams, moist soils, vegetables (Römling et al., 2005, Burns et al., 2001) and other CF patients (Hoiby and Pedersen, 1989, Armstrong et al., 2002); with certain epidemic strains more widespread within the cystic fibrosis community (Cheng et al., 1996, Frederiksen et al., 1999, Aaron et al., 2010). Although prevalence of *P. aeruginosa* is decreasing (Fig. 6.1), prolonged chronic infection in cystic fibrosis sufferers along with regular antibiotic treatments gives rise to antibiotic resistant strains (Pitt et al., 2003, Smith et al., 2006), phagocytosis resistant strains (Mahenthiralingam and Speert, 1995) and decreased immune system recognition (Zhang et al., 2005).

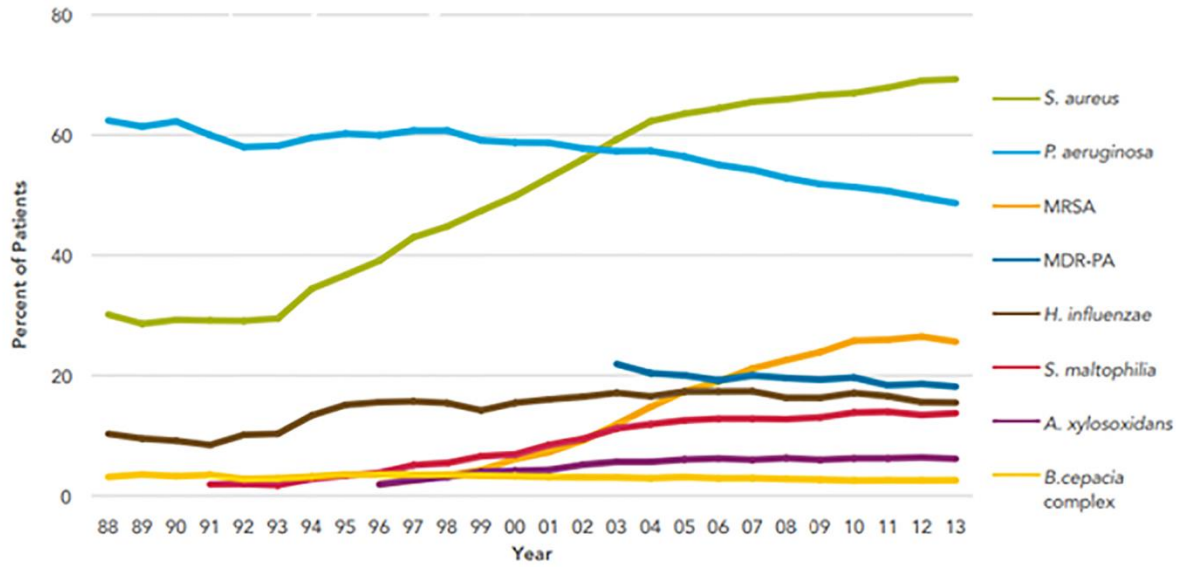


Fig. 6.1: Data from The Cyctic Fibrosis Foundation tracking the percentage of cystic fibrosis sufferers in the US infected by a variety of bacterial species between 1988 and 2013. *P. aeruginosa* infections are shown in light blue. Figure taken from the Cystic Fibrosis Foundation Patient Regsistry annual data report 2013. Cyctic Fibrosis Foundation, Bethesda, MD

P. aeruginosa infection is also prevalent in chronic wounds such as diabetic foot ulcers and burn wounds. A chronic wound is trapped in a persistent inflammatory stage with excess neutrophil proteases causing further injury and preventing proliferation of skin cells and remodelling to

properly heal wounds (Yager and Nwomeh, 1999, Diegelmann, 2003). These chronic wounds present a favourable entry environment for bacteria to colonise and proliferate; with bacteria primarily originating from the skin and environment to form polymicrobial colonies. Of these bacteria, *P. aeruginosa* are frequently isolated (Davies et al., 2004, Gjodsbol et al., 2006, Dowd et al., 2008), with wounds found to be larger and presents delayed healing compared to *P. aeruginosa* free wounds (Gjodsbol et al., 2006, Fazli et al., 2011, Madsen et al., 1996).

P. aeruginosa is also a major factor in ear infections leading to Otitis media, a common childhood infection presenting with inflammation of the middle ear; which can lead to conductive hearing loss, impeded speech and social skill development (Verhoeff et al., 2006). It is thought that biofilm formation plays a key role in the development of Otitis media, with a recent study identifying mucosal biofilms in 92% of samples tested using confocal laser scanning microscopy (Hall-Stoodley et al., 2006). The prevalence of *P. aeruginosa* in hospitals and its propensity to form biofilms means *P. aeruginosa* infection via medical devices have a significant clinical impact on morbidity and mortality as well as monetary cost (Tolker-Nielsen, 2014). Hospital acquired infections are estimated to occur in around 1.75 million patients (175,000 deaths) in Europe and 2 million infections (90,000 deaths) in the US per year (Guggenbichler et al., 2011). Of these infections, *P. aeruginosa* accounts for 10-20% of nosocomial infections (Ramos et al., 2013) and, by extension, 17,500-35,000 deaths in Europe and 9000-18,000 deaths in the US per year. Following insertion, medical devices become coated with platelets, tissue and plasma proteins; these provide a surface for microbial colonisation and biofilm formation (Francois et al., 2000). Following inoculation a thick biofilm can form within 24 hours (Chambless et al., 2006). The widespread use of medical devices means *P. aeruginosa* infection is seen in intravascular (Raad, 1998) and urinary catheters (Stickler et al., 1998), prosthetic prostheses (Martínez-Pastor et al., 2009) and breast, penile and cochlear implants (Weichman et al., 2013, Vaidyanathan et al., 2009, Germiller et al., 2005). Although the above *P. aeruginosa* infections are varied, with infection occurring in a wide range of tissues, all share a common virulence factor; the formation of biofilms.

6.1.2. Biofilms

The majority of bacteria are able to exist as either single planktonic cells or surface-anchored aggregates called biofilms. These heterogeneous bacterial communities allow bacteria to form protective microenvironments to protect against limited nutrients, dessication, toxic chemicals and fluctuations in pH and temperature (Hall-Stoodley and Stoodley, 2005). Biofilms present a significant stumbling block in treatment of many bacterial infections, providing protection against antimicrobials and the host immune system to allow persistent and chronic bacterial infections (Costerton et al., 1999). As a result, various anti-biofilm strategies have been developed to allow more effective treatment (Sommer et al., 2013, Caly et al., 2014), but many infections remain difficult to treat. Biofilm formation occurs over various stages, including attachment (reversible and irreversible), where bacteria adhere first to a surface and then to one another; followed by formation of the biofilm (maturation-1 and -2), by producing an extracellular matrix, and finally biofilm dispersal where the extracellular matrix is broken down and bacterial adhesion is reduced to facilitate emigration to new infection sites (Tolker-Nielsen, 2014, Sauer et al., 2002). Each stage involves the complex interplay between various genes and proteins to regulate biofilm progression.

6.1.2.1. Initiation of Biofilm Formation

Attachment and aggregation of bacterial cells to form a biofilm under different conditions requires a combination of several key components including the flagellum, type IV pili (twitching motility), Chaperone/usher protein (Cup) fimbria, extracellular DNA and *Ps1* polysaccharide (Tolker-Nielsen, 2014). Mutant strains of *P. aeruginosa* deficient in flagellar-mediated motility were unable to form a monolayer on abiotic surfaces (O'Toole and Kolter, 1998, Sauer et al., 2002), while type IV pili mutant strains were able to form a monolayer but not microcolonies observed in wild type strains (O'Toole and Kolter, 1998, Chiang and Burrows, 2003). In a similar vein to the type IV pili, Cup fimbria adhesins mediate attachment under different environmental conditions (Vallet et al., 2001) and have been found to be essential in upper respiratory tract infection by *Bordetella pertussis*

(Kimura et al., 1990). Initially extracellular DNA was assumed to be the product of lysed cells as opposed to a deliberate mechanism to aid biofilm formation, however, addition of DNase to cultures has been found to strongly inhibit biofilm formation in *P. aeruginosa* at the early stages of development (Whitchurch et al., 2002). Ps1, a mannose rich polysaccharide, also plays a key role in adhesion in *P. aeruginosa* strains, with *ps1A and ps1B* compromised strains demonstrating impaired biofilm initiation. Generation of an arabinose induced *ps1 P.* aeruginosa strain showed a deterioration in biofilm thickness with down regulation of *ps1* after initial attachment under arabinose conditions, suggesting the *ps1* polysaccharide plays a role in maintaining biofilm structure (Ma et al., 2006). Other polysaccharides, including the glucose-rich Pel and alginate (guluronic and mannuronic acid), are involved in adhesion and biofilm formation (Friedman and Kolter, 2004, Govan and Deretic, 1996).

6.1.2.2. Biofilm Formation and Growth

Following initial surface-attachment, biofilm formation in *P. aeruginosa* consists of non-motile cells proliferating in a fixed position to form mushroom stalks while motile cells produce the mushroom caps using type IV pili (Klausen et al., 2003). Development following initial colonisation is more dependent on environmental conditions. Citrate-grown *P. aeruginosa* in flow-chambers form a flat biofilm, with cells in initial colonies migrating when colonies reach a certain size until maturation of the biofilm; where migration stops (Klausen et al., 2003). Glucose grown biofilms on the other hand show mushroom shaped microcolonies with mushroom caps formed by motile cell populations in a flagella and type IV pili dependent manner (Barken et al., 2008, Tolker-Nielsen, 2014).

6.1.2.3. The Extracellular Matrix

Extracellular matrix components provide an essential cell-to-cell scaffold for biofilms to allow attachment, cell-to-cell interactions and protection from antimicrobials (Ma et al., 2006, Yang et al., 2009, Friedman and Kolter, 2004). *P. aeruginosa's* biofilm consists of a mix of polysaccharides

(Pel, Ps1 and alginate), proteins (type IV pili, Cup, CrdA, LecB and functional amyloids in *Pseudomonas* (Fap)) and extracellular DNA. In addition to their twitching motility role, type IV pili are also thought to act as cross linkers with extracellular DNA. As eluded to earlier, the Cup fimbria (CupA-D) are involved in attachment and formation of biofilms. CrdA, a large surface protein, and CdrB, a putative outer membrane transporter, work in combination with Ps1 to cause auto-aggregation and crosslinking of Ps1 polymers; as well as tethering Ps1 to the cell wall of *P. aeruginosa* (Borlee et al., 2010). The 6 gene *fapA-F* operon expresses functional amyloids, self-assembling protein monomers that form amyloid fibrils (Tycko, 2004). The exact function of Fap remains unknown but it appears to play a role in biofilm formation, with over expression producing "extraordinarily huge microcolonies" (Dueholm et al., 2013). The importance of extracellular DNA and polysaccharides was expanded on earlier and doesn't bear repeating.

6.1.2.4. Quorum Sensing

A key feature of biofilms is the ability for single cells to work together through cell-to-cell signalling, also referred to as quorum-sensing. Individual cell expression of quorum-sensing signalling molecules allows the colony to regulate gene expression relative to cell density, including virulence factors and regulation of the biofilm itself (Passador et al., 1993). *P. aeruginosa* quorum-sensing is coordinated through three related systems, the Las, Rhl and Pqs systems which employ N-(3-oxodoecanoyl)-L-homonoserine lactone, N-butanoyl homoserine lactone and 2-heptyl-3-hydroxy-4-quinone, respectively as signalling molecules (Juhas et al., 2005). These three pathways work in a hierarchical manner with the Las system affecting expression of Rhl system mediated expression while the Pqs system operates between the Las and Rhl systems (Tolker-Nielsen, 2014). Quorum sensing has been found to regulate several hundred genes in *P. aeruginosa* for biofilms not expressed in planktonic cells (Hentzer et al., 2005). Mutation or inhibition of any one of these systems has been found to inhibit cap formation in flow-chamber experiments (Tolker-Nielsen, 2014); in addition to this, the nutritional environment has been found to affect quorum-sensing

mutant strains with varying carbon sources causing either defects in biofilm formation or a wild-type phenotype (Shrout et al., 2006). Quorum-sensing has also been found to regulate extracellular DNA expression, *pel*, Cup, LecA and LecB expression and nitric oxide removal in nitrate respiration (Tolker-Nielsen, 2014).

6.1.2.5. Biofilm Dispersal

Dispersal of biofilms is equally important to allow bacteria to migrate and infect new areas. Emigration can be triggered by a variety of external factors such as carbon or oxygen starvation (Gjermansen et al., 2005, Thormann et al., 2006). A key component in the regulation of biofilm dispersal is the widely conserved secondary messenger, Bis-(3'-5') cyclic di-guanylate (CdG). Depletion of carbon has been found to activate the biofilm dispersion locus (*bdlA)* in *P. aeruginosa*, which produces BdlA; a chemotaxis regulator which reduces the intracellular levels of CdG via a signalling cascade (Morgan et al., 2006). Other proteins involved in biofilm dispersal include RbdA, a GGDEF and EAL domain phosphodiesterase (PDE) with a PAS low-oxygen sensing domain, which negatively regulates the extracellular matrix polysaccharides Pel and Psl and upregulates bacterial motility (An et al., 2010). EAL domains are PDEs which degrade CdG to regulate a variety of cellular processes. Another CdG level altering protein involved in biofilm dispersal is DipA, an EAL domain containing protein found to prevent nutrient-induced dispersion by glutamate, nitric oxide, ammonium chloride and mercury chloride when inactivated (Roy et al., 2012). In addition to CdG, *cis*-2-deconoic acid has been found to play a key role in biofilm dispersion. *Cis*-2-deconoic is produced continuously by *P. aeruginosa*, but when the rate of production exceeds the rate of diffusion of *cis*-2-deconoic acid away from the microcolonies biofilm dispersal is activated (Davies and Marques, 2009).

6.1.3. Bis-(3’-5’) cyclic di-guanylate

Bis-(3’-5’) cyclic di-guanylate (CdG) is an evolutionarily conserved universal second messenger in eubacteria involved in regulation of a wide variety of functions including developmental transitions, adhesion, motility synthesis of virulence factors and the formation of biofilms (Schirmer and Jenal, 2009, Boyd and O'Toole, 2012, Caly et al., 2014, Romling et al., 2013). CdG was first discovered 25 years ago in *Acetobacter xylinium* as an allosteric regulator of cellulose synthase (Ross et al., 1987, Romling et al., 2013). Intracellular levels of CdG are regulated by synthesis and degradation through diguanylate cyclases (DGCs) and phosphodiesterases (PDEs), respectively (Fig. 6.2). CdG specific DGCs are identified by their conserved GGDEF sequence and are responsible for conversion of two GTP molecules to CdG. In a similar manner, PDEs responsible for degradation of CdG to 5’-phosphoguanylyl-(3’-5’)-guanosine (pGpG) are identified by their conserved EAL sequence (Ryan et al., 2006a, Hengge, 2009); a type II CdG-specific PDE known as a HD-GYP has also been shown to possess PDE activity (Bellini et al., 2014). GGDEF, HD-GYP and EAL domain proteins will be discussed in more detail in the following section.

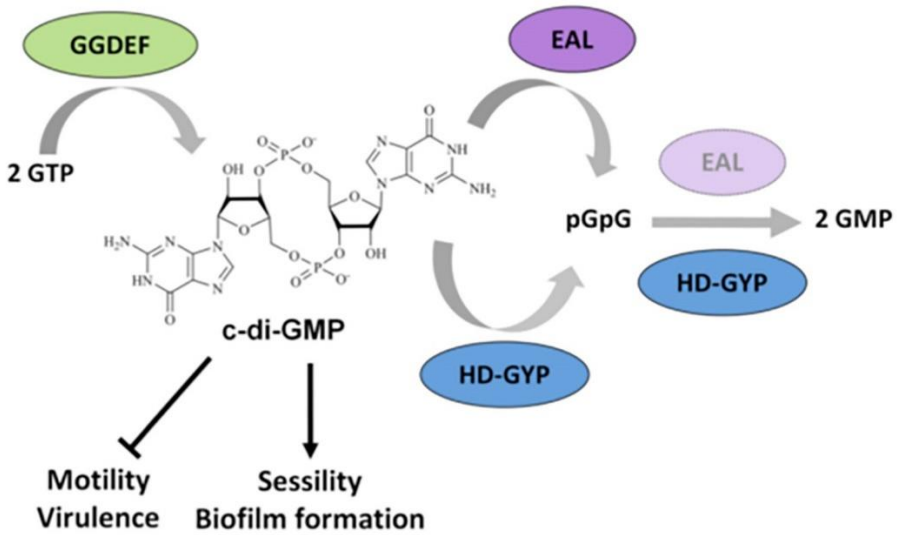


Fig. 6.2: Degradation and synthesis of cyclic di-GMP (CdG). GGDEF diguanylate cyclases synthesise CdG from two GTP molecules while CdG is converted to GMP via the intermediate pGpG by EAL and HD-GYP phosphodiesterases. Conversion to GMP by EAL domain proteins is a slower process. High CdG promotes sessility and biofilm formation while low CdG promotes motility and virulence. Figure adapted from Caly et al. (2014).

The exact mechanisms by which the changing levels in the intracellular pool of CdG affects changes in bacteria are still poorly understood. It has been suggested that environmental cues may be interpreted by differing affinity constants between CdG receptors/effectors to determine activation or repression of functions (Ryan et al., 2006b). This has been demonstrated in *Salmonella typherium* where YcgR and BcsA, two CdG sensing PilZ domain proteins, have a 43-fold difference in CdG affinity; with slight increases in CdG activating YcgR to inhibit motility and higher intracellular levels required to activate cellulose production through BcsA (Pultz et al., 2012). The 8 PilZ containing proteins in *P. aeruginosa* (*PA*3353, *PA*2989, *PA*0012, *PA*4324, *PA*2799, *PA*4608, *PA*3542 and *PA*2960) showed an even larger difference between maximum and minimum affinities with a 145-fold difference (Christen et al., 2010, Pultz et al., 2012). An alternative, but not mutually exclusive theory, is that CdG signalling consists of different systems related to specific tasks based on discrete pools of CdG; as demonstrated by regulation of entry into the S phase of the cell cycle through PopA following up regulation of CdG levels by DgcB and PdeA in *Caulobacter crescentus* (Abel et al., 2011).

Despite the controversy surrounding CdGs mechanism of regulation it is evident that CdG is able to interact with a large number of cellular targets. To date CdG has been found to interact with PilZ domains (a 100 amino acid repeat found in various proteins) (Amikam and Galperin, 2006), transcription factors including FleQ in *P. aeruginosa*, VpsT in *Vibrio cholera,* Clp in *Xanthomonas* species, Bcam1349 in *Burkholderia cenocepacia* and LtmA in *Mycobacterium smegmatis* (Caly et al., 2014). CdG also interacts with riboswitches, part of the 5' untranslated region (UTR) of mRNAs, via GEMM elements found in the UTR of PDE and DGC genes; as demonstrated by activation of the *tfoX* gene in *V. cholera* through CdG binding (Sudarsan et al., 2008). Although GGDEF domains and HD-GYP/EAL domains main function involves the formation and degradation of CdG, respectively; there are a certain population of enzymatically inactive variants of these proteins that are thought to act as sensory domains (Caly et al., 2014).

6.1.4. CdG Synthesis by Diguanylate Cyslases

CdG specific DGC domain (GGDEF) activity was first identified via PleD of *C. crescentus*, a multidomain protein consisting of N-terminal CheY-like receiver domains and a C-terminal domain later identified as a GGDEF domain (Paul et al., 2004, Chan et al., 2004). Activation of PleD by phosphorylation correlated with DGC activity and the conversion of two GTP molecules to CdG (Paul et al., 2004). Following on from this, six more GGDEF domain proteins from *Thermatoga, Deinococcus-Thermus, Cynobacteria, Spirochetes and Proterobacteria* (α and γ) were identified and confirmed to possess DGC activity for CdG in the presence of Mg^{2+} (Ryjenkov et al., 2005). GGDEF domains adopt a conserved five stranded antiparallel β-sheet surrounded by helices (Fig. 6.3), similar to the catalytic core of adenylate cyclase (Tesmer et al., 1999, Chan et al., 2004). The GG(D/E)EF motif is located in the connecting loop between β2 and β3 (Fig 6.3b). This GGDEF motif forms the active "A site" where a single molecule of GTP binds per monomer (Caly et al., 2014). A second conserved sequence (RxxD) (Chan et al., 2004), later identified as an inhibitory site through structural analysis of CdG-bound PleD and WspR, was found to be involved in feedback inhibition of GGDEF domains by preventing formation of the enzymatically active DGC domain dimer (Wassmann et al., 2007, Schneider, 2008).

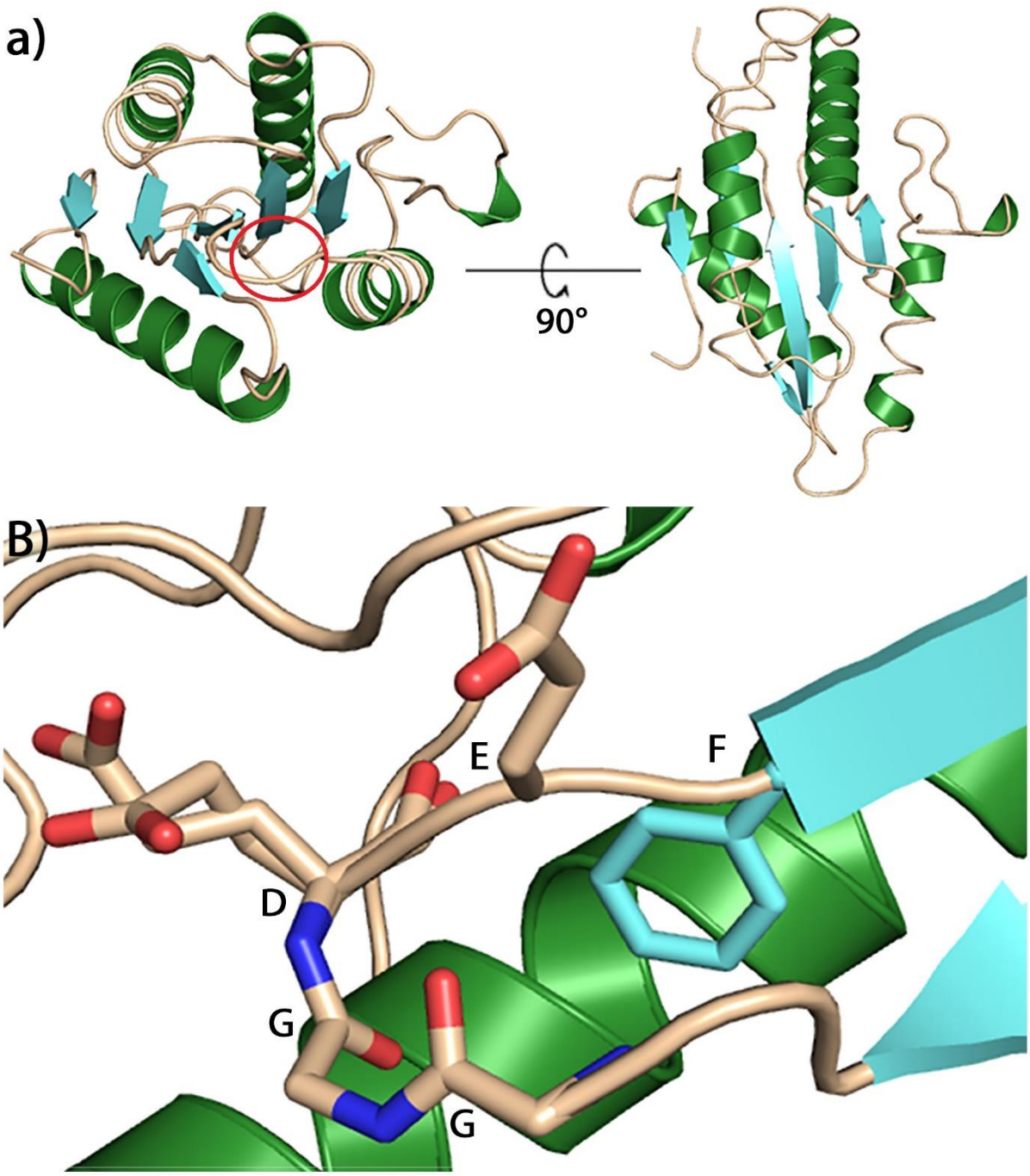


Fig. 6.3: The crystal structure of *PA*4332-GGDEF. a) Structure of *PA*4332-GGDEF monomer shown in cartoon representation with α-helices coloured green, β-sheets in blue and loops in wheat. The GGDEF domain is circled in red. b) A detailed view of the conserved GGDEF sequence in the active "A site" with side chains shown as sticks and coloured by elements.

Studies of GGDEF domains in solution and through crystal structures suggest DGC activity is mediated by formation of a dimer. Both PleD and WspR demonstrate dynamic quaternary structure changes to facilitate inhibition of DGC activity in the presence of CdG. PleD consists of a receiver

domain, adaptor domain and DCG domain depicted in red, yellow and green, respectively in figure 6.4; all connected by flexible linker domains. Following phosphorylation of the receiver domain (Fig. 6.4(2)) the PleD receiver and adaptor domains dimerise (Fig. 6.4(3)), facilitating DGC domain dimerisation and formation of CdG from two bound GTP molecules in the "A site" of each monomer (Fig. 6.4(4)) (Wassmann et al., 2007). Inhibition of PleD can occur by two distinct interactions between CdG and inhibitory sites on the DGC and adaptor domains; i) dimeric CdG molecules bind the primary "I site" of the DGC domain (I_P) and the inhibitory site on the adaptor domain (Fig. 6.4(5)). ii) Following phosphorylation of the receiver domain, dimeric CdG binds the primary and secondary "I sites" of the DGC domain to produce a non-productive dimer (Fig. 6.4(6)) (Wassmann et al., 2007). WspR undergoes an equally sophisticated mode of regulation by dimerisation, elegantly demonstrated by crystal structures of each oligomeric state (Fig 6.5). Active WspR forms a compact parallel coiled coil dimer mediated by the CheY-homology domains to facilitate DCG domain dimerisation (Fig. 6.5.a). Inhibited WspR first takes the form of an antiparallel tetrameric coiled coil, facilitated by a dimeric CdG clamp which binds the DCG domain "I site" to stabilise the conformation of inactive dimers and prevent dimerisation of the DGC domains (Fig. 6.5b). Once stabilised, inhibited dimers can detach from the tetramer (Fig. 6.5c) (Schneider, 2008). Reactivation of WspR is achieved by PDE mediated removal of CdG to form the active parallel dimer structure in Figure 6.5a.

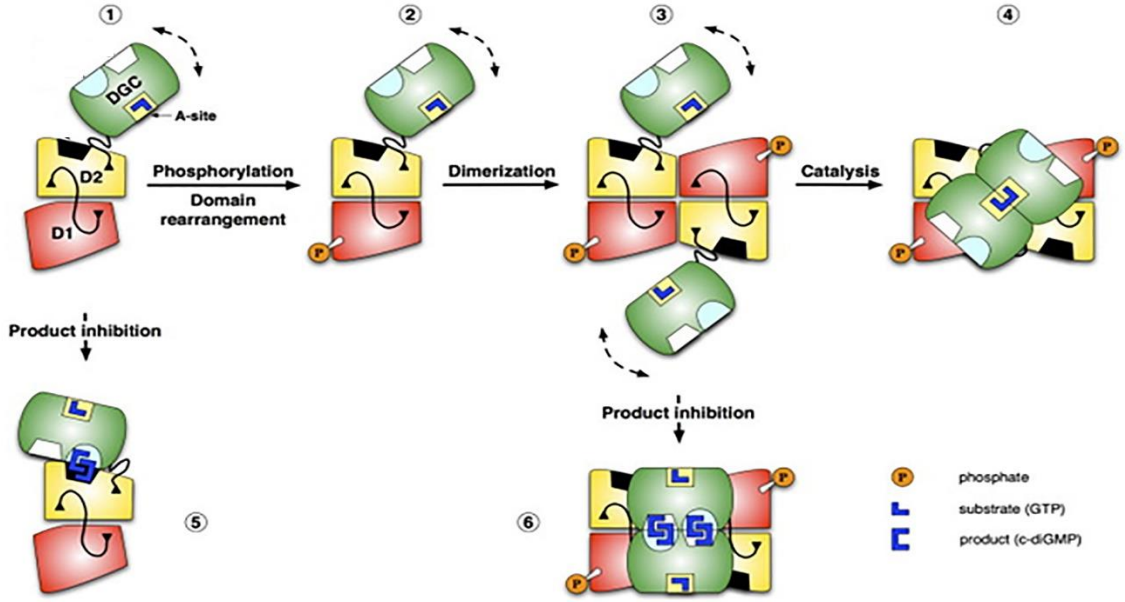


Fig. 6.4: A mechanistic model of PleD regulations by CdG and dimerisation. PleD consists of a diguanylatecyclase domain (green), a receiver domain (D1, red) and adaptor domain (D2, yellow) connected by flexible linkers (black). 1-4 depict PleD activation by dimerisation with monomeric PleD (1) undergoing phosphorylation (orange) to bring about a domain rearrangement (2) leading to dimerisation of D1 and D2 domains (3) to facilitate dimerisation of the GTP (blue) bound DGC monomers to form cyclic di-GMP (CdG). 5-6 show two methods of inhibition. Dimeric CdG molecules can bind the D2 inhibition site (black) and primary inhibition site of the DGC domain (cyan) (5). Alternatively Dimeric CdG binds both DGC monomer's primary (cyan) and secondary inhibition sites (white) (6). Both prevent effective dimerisation to form the active site and have been observed in crystal structures. Figure adapted from Wassmann et al. (2007).

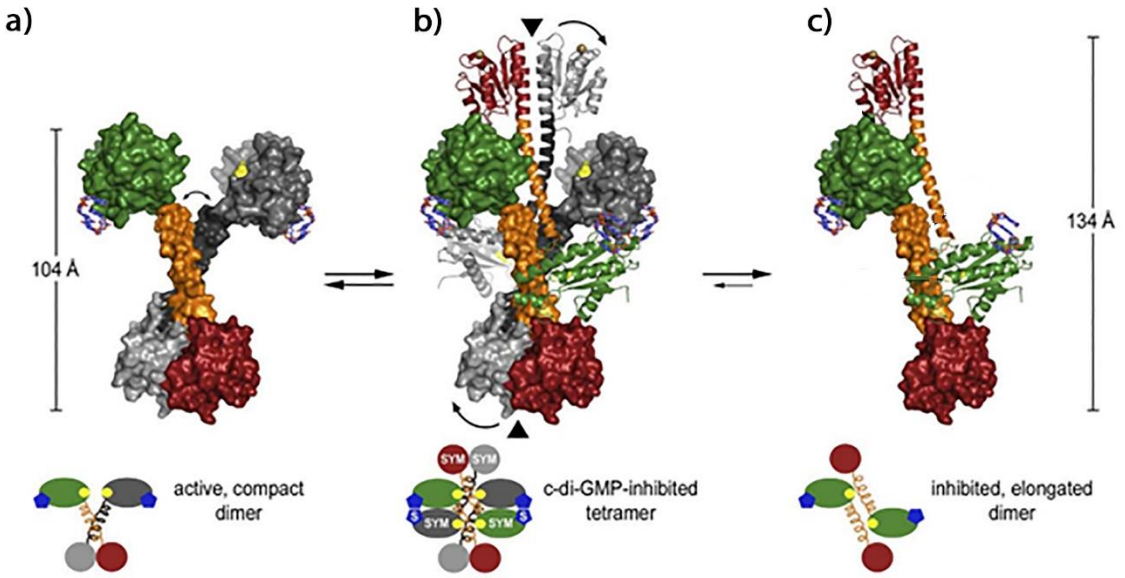


Fig. 6.5: A mechanistic model of WspR regulation with accompanying crystal structures. For all structures and cartoons CheY domains are coloured red, stalks in orange, GGDEF domains in green with a yellow active site and cyclic di-GMP (CdG) molecules in blue. a) A structural model shown in surface mode and accompanying cartoon of the parallel active compact dimer of WspR facilitate by the CheY domain. b) The inhibited WspR anti-parallel tetramer facilitated by CdG binding to GGDEF domain "I sites" with the stalks preventing dimerisation of the GGDEF domains. Domains labelled S or SYM in cartoons indicated a crystal symmetry related molecule. c) The elongated inhibited anti-parallel WspR dimer released from the tetramer form. Figure adapted from De et al. (2008)

6.1.5. CdG Degradation by Phosphodiesterases

6.1.5.1. HD-GYP Phosphodiesterases

Degradation of CdG is undertaken by two unrelated PDEs, HD-GYPs and EAL domain containing proteins (Fig. 6.2). HD-GYP domains are a variant of HD-type phosphohydrolases, defined by a C-terminal sub domain with additional conserved residues; widely distributed but less abundant than EAL PDEs (Galperin, 2005). HD-GYP domains were first identified as a metal-dependant (Mg^{2+} or Mn^{2+}) CdG specific PDE capable of hydrolysing CdG to two GTP molecules with pGpG as an intermediate (Ryan et al., 2006a). The role of HD-GYPs in regulation of biofilms was further investigated though four HD-GYP domain proteins in *V. cholera (Hammer and Bassler, 2009), BB*0374/PdeB from *Borrelia burgdorferi* (Sultan et al., 2011) and *PA*4108 and *PA*4781 from *P. aeruginosa* (Ryan et al., 2009); all of which agreed that mutation of these HD-GYP domains led to a decrease in CdG levels and reduced biofilm formation. Alanine substitution of individual residues from the GYP motif affected the degree of interaction with CdG and total alanine substitution (HD-AAA) caused almost complete inhibition. Alanine substitution of the HD motif had no effect on catalytic activity (Ryan et al., 2010). The first structure of a HD-GYP domain was the inactive *Bd*1817, which lacks the catalytic tyrosine from the HD-GYP motif (Lovering et al., 2011). *Bd*1817 contained an unidentified N-terminal domain and a C-terminal HD-GYP domain separated by a 3 helix linker. The HD-GYP domain is completely alpha helical with a binuclear metal centre with two Fe^{2+} ions octahedrally coordinated by conserved residues and bidentately coordinated by a water molecule and phosphate ion (Fig. 6.6a-b). The extended HD-GYP motif, or HEExxNGxG-P in the case of *Bd*1817, is located at the solvent exposed active site pocket and forms a novel double U-shaped turn, shown in orange in figure 6.6c-d (Lovering et al., 2011). The first structure of a catalytically active HD-GYP domain was that of *Pm*GH from *Persephonella marina*, solved in the presence of the substrate (CdG)

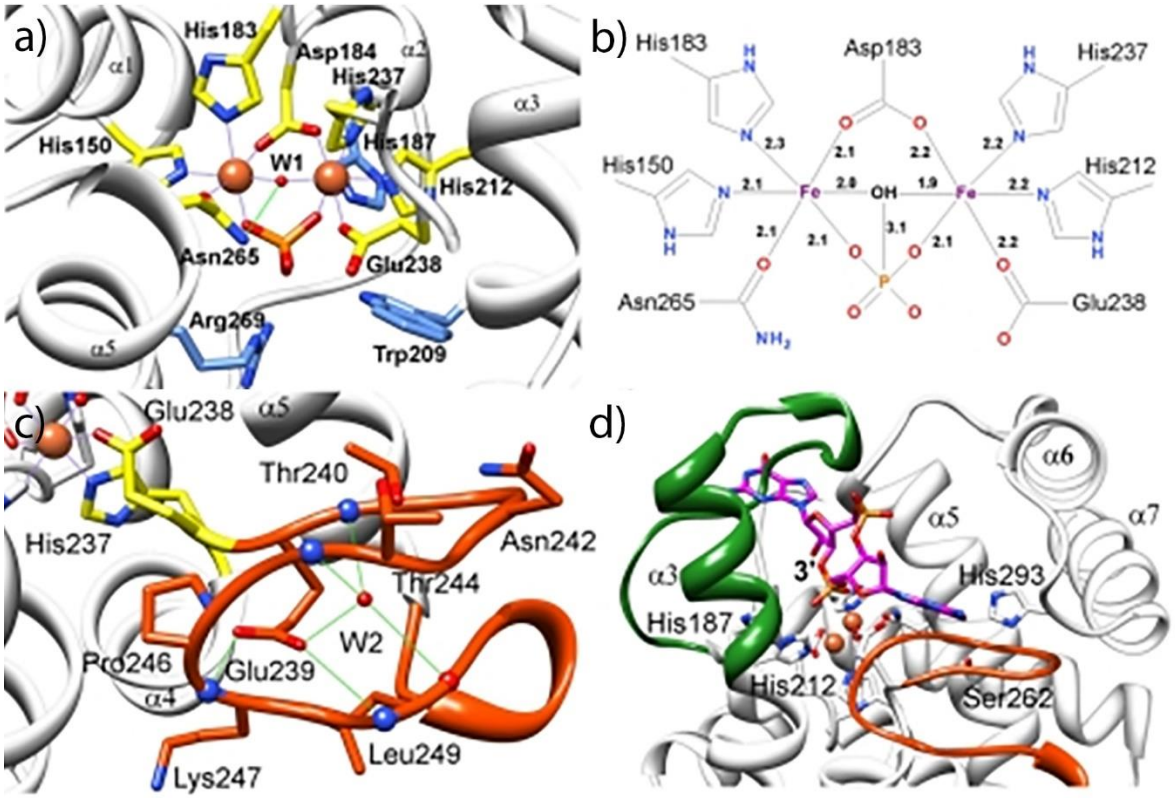


Fig. 6.6: Bd1817 HD-GYP active site metal coordination and the structure of the GYP region and possible binding site. a) A detailed view of the bi-metallic active site with metal coordinating residues and the bound phosphate shown as sticks. Fe and hydroxide ions are depicted as spheres in orange and red, respectively. Fe-protein interactions are represented by purple lines and hydroxide-protein interactions as green lines. b) Schematic of the bi-metallic active site with coordination distance measures in angstroms. c) The extended GYP motif (Residues 239-255) in orange shown as sticks and ribbons to demonstrate the novel double U-shaped turn. d). A predicted cyclic di-GMP (CdG) binding site with CdG binding modelled based on the position of the phosphate ion shown in A. The GYP motif (orange) and lid region (green) for the predicted binding pocket. Figures adapted from Lovering et al. (2011).

and final reaction product (GMP) (Bellini et al., 2014). *Pm*GH consists of an N-terminal GAF domain, 42 residue linker helix and a C-terminal HD-GYP domain, crystallised as a parallel dimer with binding mediated by the GAF domain (Fig 6.7). The structure of the overall fold of the HD-GYP domain is well conserved between *Pm*GH and *Bd*1817, as shown in figure 6.8a. A clear shift is observed in α6 and α10 to accommodate the trinuclear metal site of *Pm*GH (Fig. 6.8a). α1 and α5 of the HD domain also show a conformational change, shifting from an open conformation in *Pm*GH (red loops in figure 6.8b) to a closed conformation in *Bd*1817 mediated by hydrogen bonds (green helices in figure

6.8b). The tri and binuclear metal sites in *Pm*GH and *Bd*1817, respectively, show similar liganding residues between metal centres with most differences arising only in response to the new third metal observed in *Pm*GH (Fig. 6.8c). CdG-bound *Pm*GH suggests the trinuclear metal centre interacts directly with the scissile phosphodiester of CdG to act as a Lewis acid catalyst to activate a bridging water for nucleophilic attack (Bellini et al., 2014). Interestingly, the *cis* binding conformation of CdG observed in this HD-GYP (Fig. 6.9a) domain differs greatly from the extended conformations in EAL domains (Fig. 6.9b) which carry out the same hydrolysis reaction.

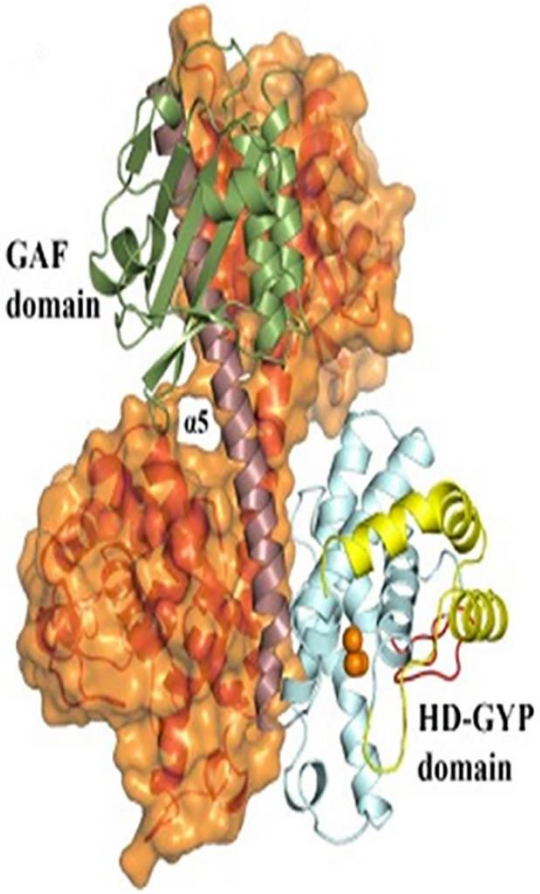


Fig. 6.7: Crystal structure of PmGH homo dimer. Molecule A is shown in cartoon form with the GAF domain coloured in green and the inter-domain dimerisation helix in purple. The HD-GYP domain is split into the core HD domain in cyan, the GYP-motif domain in red and additional helices in yellow. Molecule B is shown as a transparent surface representation (orange) with cartoon representation as in molecule A. The tri-metallic iron centre is represented by orange spheres. Figure adapted from Bellini et al. (2014)

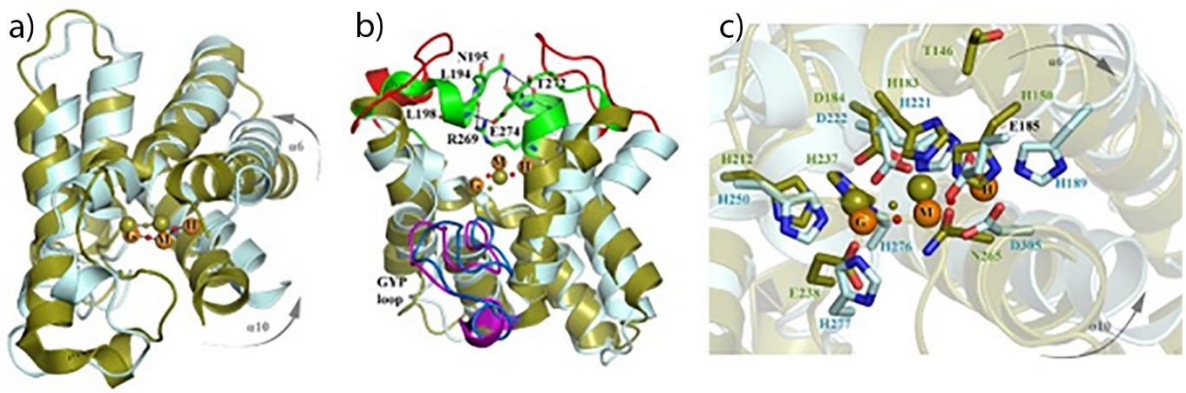


Fig. 6.8: Structural comparisons of PmGH (blue) and Bd1817 (green) HD-GYP domains. Both Structures are presented as cartoons with metal ions shown as spheres in orange and green for PmGH and Bd1817, respectively. a) Comparison of overall folds between proteins and helices involved in forming the metal binding site; labelled α6 and α10. b) Comparison of open and closed nucleotide binding pockets with open PmGH in red and closed Bd1817 in green. Hydrogen bonding Bd1817 residues are represented as sticks with bonds as black lines. c) Metal liganding residues of PmGH and Bd1817. Figure adapted from Bellini et al. (2014)

6.1.5.2. EAL Phosphodiesterases

EAL domain proteins have been extensively studied, in general, and are structurally and functionally related with a highly conserved triosephosphate isomerase (TIM)-barrel fold with the active site at the C-terminal end of the barrel. The lion's share of EAL domains possess PDE activity, with some degenerated EAL domains able to bind CdG but lack the metal coordinating residues required for hydrolysis. A large cohort of published data shows active EAL domains hydrolyse CdG in the presence of Mg^{2+} and Mn^{2+}, with peak reaction rates occurring at alkaline pH (Sundriyal et al., 2014). Of these two metals manganese-loaded EAL domains have been shown to be at least one order of magnitude faster than magnesium-loaded EALs, while calcium causes complete inhibition of PDE activity (Tchigvintsev et al., 2010, Schmidt et al., 2005, Tamayo et al., 2005, Rao et al., 2008). The first reaction mechanism proposed involved a single divalent metal bound to the active site based on the structure of *td*EAL and YkuL with CdG bound in combination with mutagenesis and pH dependent studies on RocR (Rao et al., 2008, Minasov et al., 2009). Mutagenesis studies of RocR probed the catalytic importance of fourteen conserved polar residues (Fig. 6.10) including eight acidic residues (E175, E265, D295, D298, D318, E352 and E355), two basic residues (R179 and K316)

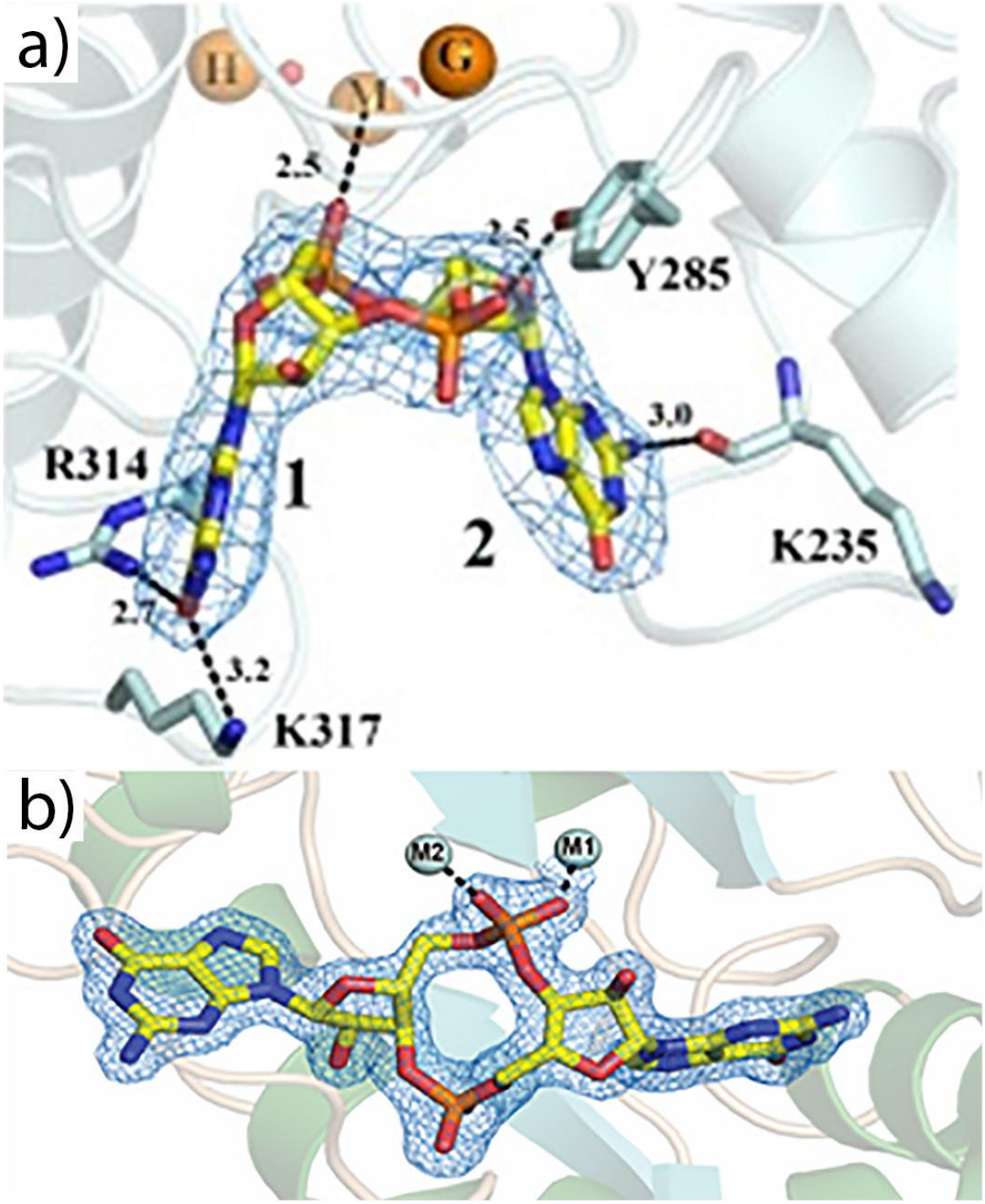


Fig. 6.9: Cyclic di-GMP (CdG) binding in PmGH and PA3825-EAL. CdG is shown as sticks coloured by atom type and iron, water and magnesium ions are shown as spheres in orange, red and cyan respectively. a) The cis binding conformation observed in PmGH with omit map density contoured to 2 σ. b) The extended conformation of CdG binding observed in PA3825-EAL with 2Fo-Fc density contoured to 1.8 σ. Figure 6.9a was adapted from Bellini et al, (2014).

and four neutral residues (Q161, Q372, N233 and T267) (Rao et al., 2008). Of these fourteen residues mutation of three residues showed a negligle to 33.5 fold reduction in the K_{cat} (R179A, T267A, D296A and E355A), suggesting minor roles in catalysis. Alanine substitution of a further three residues (Q161A, R179A and D296A) demonstrated a two- threefold increase in K_m values compared

to wild type RocR, suggesting potential roles in substrate binding (Rao et al., 2008). Mutation of the remaining seven residues (E175A, N233A, E265A, E268A, D295A, K316A and E352A) reduced activity to undetectable levels, suggesting an essential role in forming the active site. Of

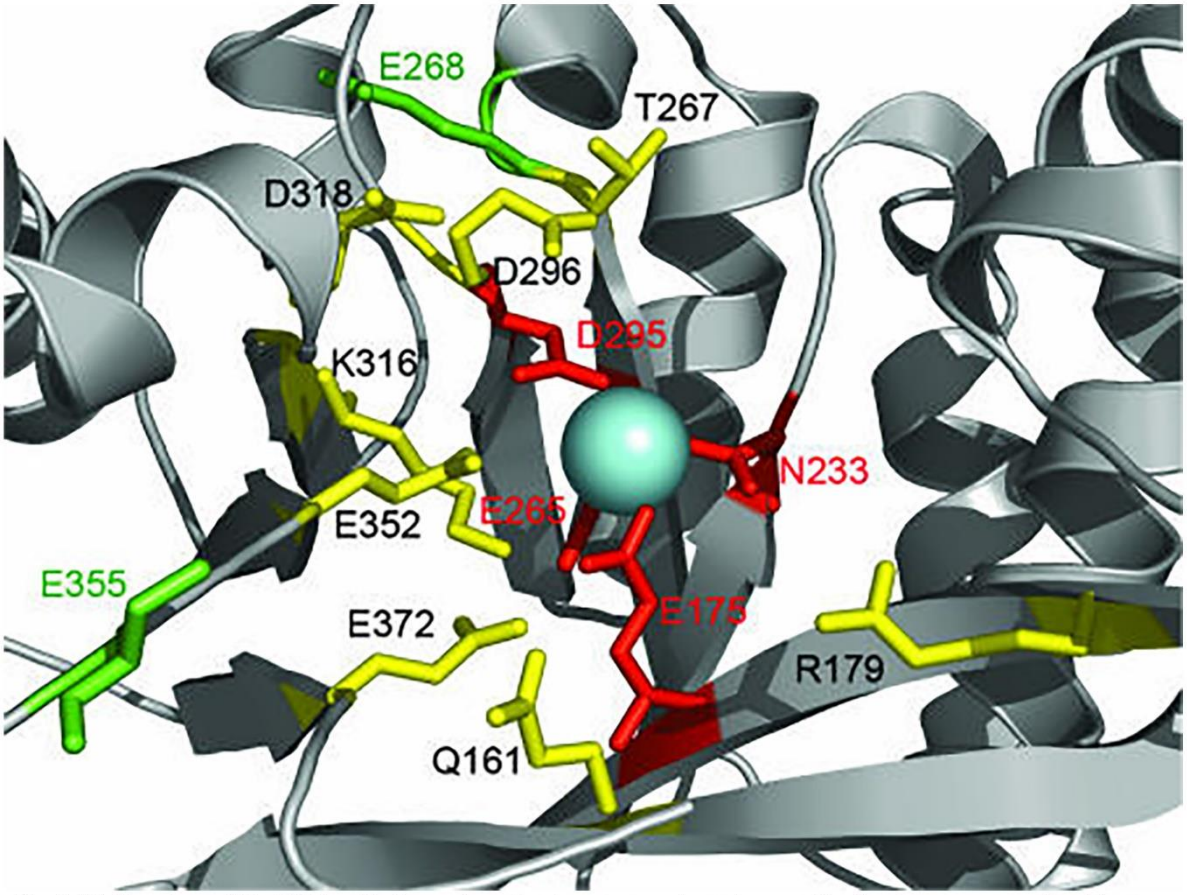


Fig. 6.10: Structure of RocR-EAL with Secondary structure as a cartoon in grey, the fourteen conserved residues as sticks and MG^{2+} ion as a cyan sphere. Metal coordinating residues are coloured red, distal residues in green and other conserved residues in yellow. Figure taken from Rao et al. (2008).

these seven, activity could be completely restored for N233A in the presence of excess Mg^{2+} (0.5 M) and partially recovered for E175A, E265A, D295A, E268A, and K316A. Only E352A showed negligible recovery in the face of excess Mg^{2+}, suggesting an essential role in catalysis later eluded to as a potential general base catalyst for deprotonation of the catalytic water for hydrolysis (Rao et al., 2008). The two metal structure of *Thiobascillus denitrificans* EAL suggested a two-metal-ion mechanism akin to polymerases and nucleases (Rao et al., 2008), however, this was dismissed based on large inter-metal distances and mutation of the second metal coordinating residues (D296 and D318) proving non-essential to function (Rao et al., 2008).

Following this intital work, the two-metal-ion catalytic mechanism was rekindled with the structure of Blrp1 from *Klebsiella pnuemoniae*, consisting of an N-terminal sensor of blue-light using FAD (BLUF) domain and C-terminal EAL domain with PDE activity (Barends et al., 2009). Blrp1 structures were solved in the presence of Mn^{2+} and CdG at a range of pH and presented with a binuclear active site of Mn^{2+} ions coordianted by conserved Asp residues (Asp302-303) on loop 6 at the C-terminal end of the TIM-barrel (Fig. 6.11) (Barends et al., 2009). Based on this structural information Barends *et al* suggest a "tight" bimetallic centre, in which two metals can activate the in-line attacking water to cleave the phosphodiester bond in CdG when they reach critical proximity. This model was based on a decreasing inter-metal distance with increasing pH in crystal structures in conjunction with *in vitro* kinetic observations in solution where reaction rates are greatly increased between pH 6 and pH 9. Structures at pH 6 and pH 9 were determined at 2.05 Å (3GFZ) and 2.55 Å (3GGO), respectively and showed a decrease in inter-metal distance of 0.4 Å and 0.6 Å in molecules A and B, respectively, with increasing pH (Fig. 6.12). However, the manganese-loaded bimetallic structure reported at an activating pH shows an intact CdG molecule in the catalytic pocket instead of the expected product, pGpG, if the protein is truly active (Barends et al., 2009). To further examine the catalytic redulation of EAL domains, two previously structurally uncharacterised EAL domains from the core genome of *P. aeruginosa*, MucR-EAL and *PA*3825-EAL, have been examined in solution and by X-ray crystallography.

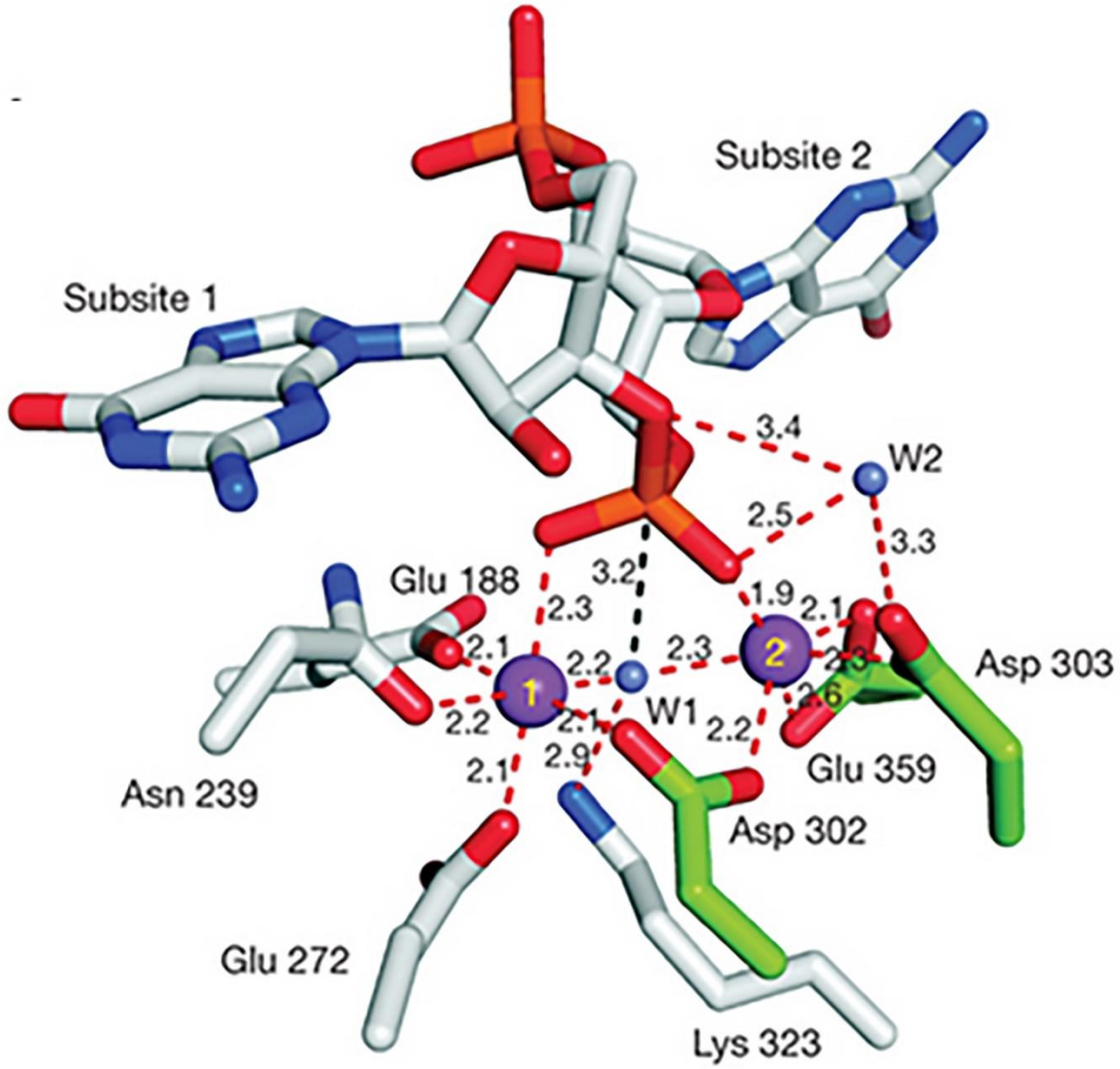


Fig. 6.11: A detailed view of the manganese-loaded Blrp1 cyclic di-GMP (CdG) bound active site. The CdG molecule and metal coordinating residues are represented in stick form with manganese and water molecules as purple and blue spheres, respectively. Distances are shown in angstroms with red dashed lines depecting protein-metal, metal-water and metal-CdG interactions and hydrolytic attack interactions as black dashed lines. Highly conserved Asp residues from loop 6 are coloured in green. Figure adapted from Barends et al. (2009).

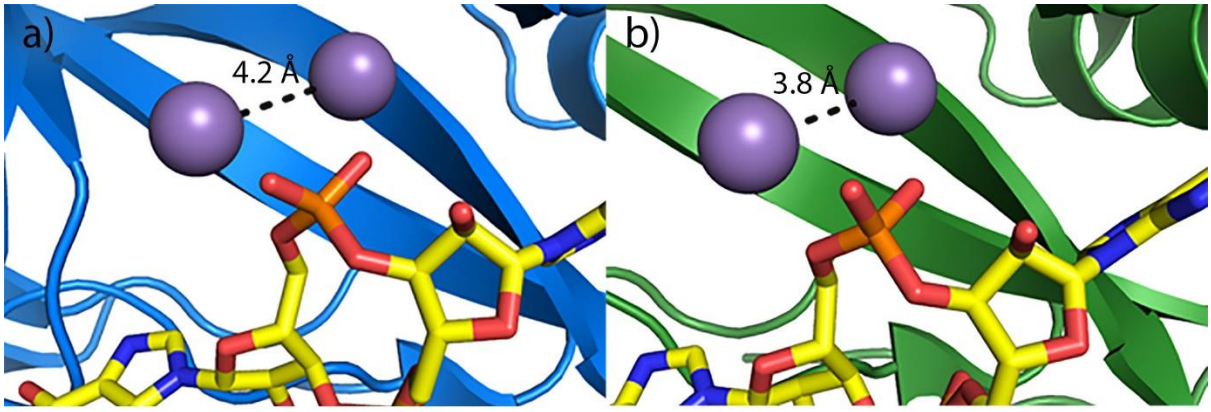


Fig. 6.12: Structure of manganese loaded Blrp1 at inactive and active pH. CdG molecules are shown as sticks, manganese ions as spheres and inter-metal distances as dashed black lines. a) Inactive Blrp1 crystallised at pH 6 with an inter-metal distance of 4.2 Å (3GFZ). b) Active Blrp1 crystallised at pH 9 with an inter-metal distance of 3.8 Å (3GG0)

PA1727 (MucR) is a putative membrane associated protein comprised of an active DGC domain (GGDEF), inactive PDE domain (EAL) and a proposed oxygen, CO or NO sensing MHYT domain (Galperin et al., 2001), involved in alginate biosynthesis and biofilm formation (Hay et al., 2009). Knockout studies of MucR in PDO300 *P. aeruginosa* generated a non-mucoid phenotype, increased swarming motility and a 38-fold reduction in alginate production; while upregulation of MucR results in a 7-fold increase in alginate production, wrinkly colony morphology, increased pellicle formation, auto-aggregation, and generation of structured biofilms (Hay et al., 2009). MucR regulation of alginate synthesis and biofilm formation is linked to activation of Alg44 by increasing localised CdG concentration through PDE activity. Despite biochemical evidence that MucR is an inactive PDE, MucR posesses all the necessary conserved residues for catalysis; making the mechanism by which MucR PDE activity is inhibitied a mystery. *PA*3825-EAL is a two domain cytoplasmic protein consisting of an N-terminal CSS domain and an active C-terminal EAL domain. CSS domains are a poorly understood receptors predicted to have a simialr fold to extracellular receptors of bacterial dimeric histidine kinases. To date they remain uncharacterised by both biochemcial and structural techniques.

In this chapter, to further elucidate the metal-centred mechanism of the EAL domain, the molecular biology, crystallisation and crystal structures of *PA*3825-EAL in its metal-free state and loaded with either Ca^{2+}, Mg^{2+} or Mn^{2+} ions, with and without CdG as well as complexed with the

reaction product pGpG are presented. The latter structure reveals a novel third metal binding site coordinated to the free oxygen atoms formed by hydrolysis of the scissile phosphodiester bond of CdG. The position of this new metal site suggests a possible catalytic role in activation of the transition state for hydrolysis. Furthermore, a combination of kinetic studies and solution state oligomerisation analyses have also been performed on *PA*3825-EAL and MucR-EAL recombinant proteins that, together with the crystal structures, including that of CdG-bound MucR_EAL, shown here for the first time, suggests a specific type of dimerisation in which a long helix (helix 8) may play a role in the inhbition of the EAL Domains.

6.2. Methods

6.2.1. Cloning, Expression and Purification

Six DNA fragments of varying lengths encoding the EAL domain of *PA*3825 (residues 255-517) were amplified by PCR from the *Pseudomonas aeruginosa* strain PAO1 genome before subcloning into pOPINF vectors using In-Fusion™ cloning with a cleavable N-terminal His_6-tag. Plasmids were transformed into Lemo21 (DE3) *E. coli* cells in the presence of 50 µg/ml ampicillin. Cultures were grown to an OD_{600} of 0.4-0.6 and induced with 0.15 mM IPTG at 18°C for 16 hours. Cultures were spun at 5000 rpm (Sorvall SLC6000 rotor) for 10 minutes at 4°C and resuspended in buffer A (0.1 M Tris-HCl pH 8.0, 0.5 M NaCl, 1 mM TCEP, 20 mM imidazole and 2% glycerol) plus 10 mM $MgCl_2$, 10 µg/ml DNAse and 300 µg of lysozyme per gram of pellet. Cells were lysed by sonication in two cycles of 45 and 30 seconds and clarified by centrifugation at 16,000 rpm at 4°C for 15 minutes (Sorvall SS34 rotor). Cleared lysate was loaded onto a Ni-NTA affinity column prequilibrated in buffer A with affinity chromatography performed on an AKTA purifier (GE Healthcare). The N-terminal His_6-tag was cleaved overnight at 4°C and the cleaved tag and other impurities removed by a second round of Ni affinity chromatography. Untagged *PA*3825-EAL containing fractions, as judged by absorbance at 280 nm, were pooled to a final volume of 6 ml and

filtered with a 0.22 µm syringe filter before loading onto a Superdex 200 16/60 gel filtration column prequillibrated in size exclusion chromatography (SEC) buffer (20 mM Tris-HCl, pH 8, 300 mM NaCl, 1mM TCEP) and eluted over one column volume. Purified MucR protein was provided by the Tews group in Southampton in 50 mM HEPES pH 7, 300 mM NaCl, 2mM $MgCl_2$ and 2 mM β-mercaptoethanol.

6.2.2. Crystallisation

Purified *PA*3825-EAL was concentrated to 15-20 mg/ml as quantified by absorbance at 280 nm for all crystallisation experiments. Initial crystal trials were set up using the sitting drop vapour diffusion method with a Cartesian crystal robot in 200 nL drops consisting of protein and reservoir solution from Wizard I + II, Wizard III + IV, JCSG-plus HT-96, Morpheus, PEG/ION, and Index commercial crystallisation screens at a ratio of 1:1. Trays were incubated at 4°C and 18°C. Optimisable crystal hits were found in 3 conditions from the Wizard I + II screen as shown in figure 6.13. Small rhombus shaped crystals were found in condition D4 (10% isopropanol, 0.1 M Mes pH 6.0 and 0.2 M calcium acetate) and large bipyramidal crystals in condition E4 (2.0 M ammonium sulphate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 0.2 M sodium chloride) and G11 (0.8 M Sodium phosphate monobasic, 1.2 M potassium phosphate dibasic and 0.1 M sodium acetate pH 4.5) as shown in figure 6.13a, b and c, respectively. G11 and E4 screens were optimised using a Hamilton Star Lab liquid Handler to produce precipitant gradients and pH screens with crystals grown in 500 nL sitting drops set up using a TPP mosquito LCP crystallisation robot. The best primitive tetragonal crystals (G11) grew at 4°C in 0.8 M sodium phosphate monobasic, 1.2 M potassium phosphate dibasic and 0.1 M sodium acetate pH 4.5 (final pH measured as pH 7.1), whereas the best centred monoclinic crystals (E4) were grown at 4°C in 6-11% iso-propanol, 0.1 M Mes pH 6.5, 0.1 M sodium acetate pH 4.5 and 0.2 M calcium acetate (final pH measured as pH 5.2).

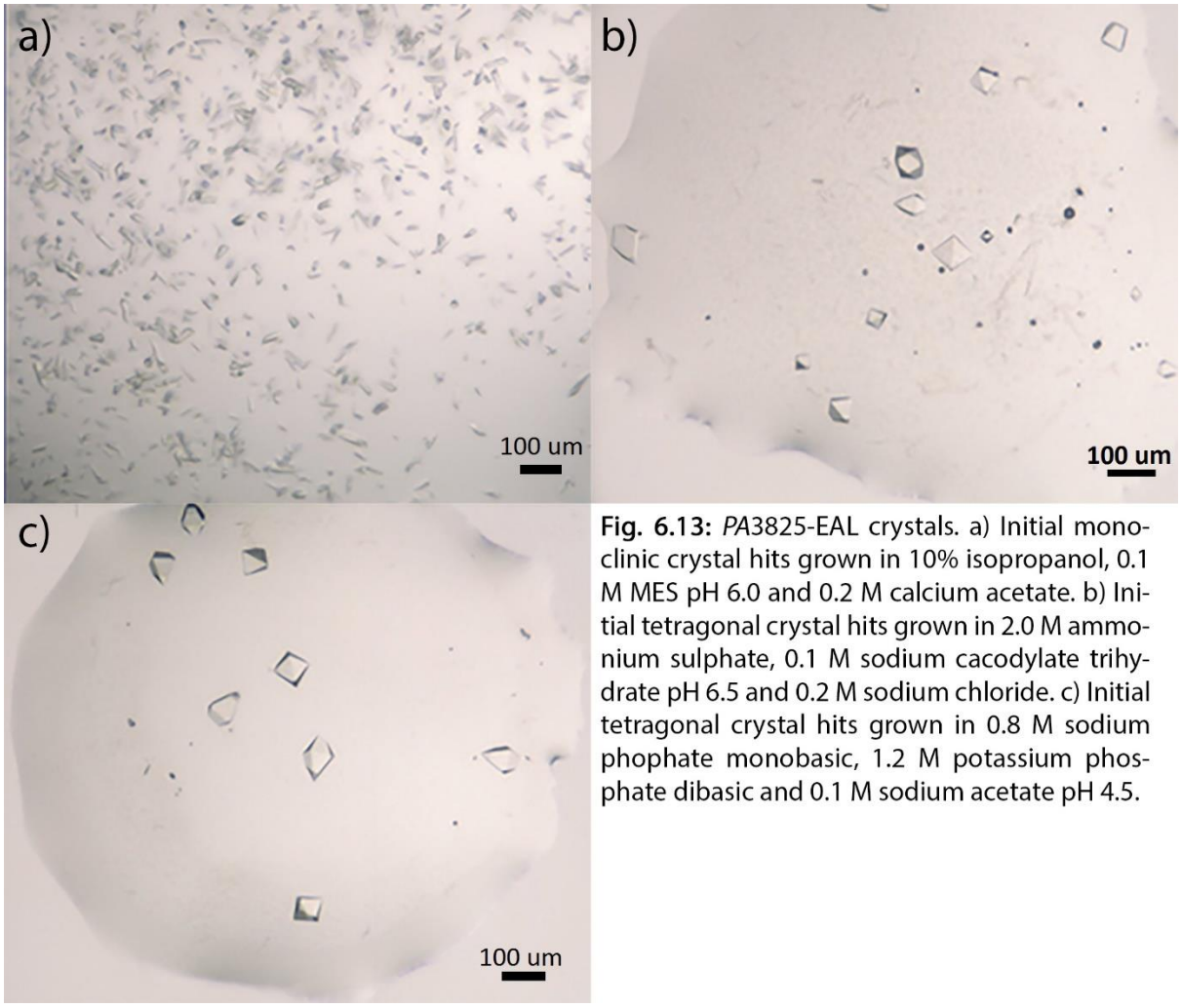


Fig. 6.13: *PA*3825-EAL crystals. a) Initial monoclinic crystal hits grown in 10% isopropanol, 0.1 M MES pH 6.0 and 0.2 M calcium acetate. b) Initial tetragonal crystal hits grown in 2.0 M ammonium sulphate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 0.2 M sodium chloride. c) Initial tetragonal crystal hits grown in 0.8 M sodium phophate monobasic, 1.2 M potassium phosphate dibasic and 0.1 M sodium acetate pH 4.5.

Co-crystallisation of tetragonal *PA*3825-EAL crystals with magnesium or manganese was performed by adding 0.2 M $MgCl_2$ or $MnCl_2$ to optimised reservoir solution prior to dispensing. Co-crystallisation of tetragonal *PA*3825-EAL crystals with CdG was achieved by adding 10 mM CdG directly to the protein sample. Soaking experiments for tetragonal *PA*3825-EAL crystals were performed over several days at 4°C with 10 mM CdG in optimised reservoir solution. Similar soaking experiments were performed with monoclinic *PA*3825-EAL crystals over one hour at 20°C. Cryoprotection was achieved by minimising surrounding solvent for primitive tetragonal crystals and adding 25% ethylene glycol to the crystallization buffer for monoclinic crystals.

6.2.3. Data Collection

Single crystal diffraction datasets for different metal and ligand co-crystallised/ soaked crystals were collected at 100 K over multiple visits on beamlines I02, I04 and I04-1 at Diamond Light Source, UK. Manganese-apo and manganese-CdG complex crystals were collected at the manganese edge on I02 at Diamond. Data collection parameters are summarised in table 6.1.

Table 6.1: Data collection parameters for *PA*3825-EAL protein crystals in the presence of bound metals (Ca^{2+}, Mg^{2+}, Mn^{2+} and Na^{+}) and ligands (apo, CdG and pGpG).

Crystal	Apo-Apo	Ca-Apo	Ca-CdG	Mg-CdG	Mn-CdG	Mn-pGpG
Wavelength (Å)	0.9763	0.9795	0.9200	0.9795	0.9795	1.8785
Beamline	I03	I04	I04-1	I02	I04	I02
Beam Size (µm)	50 x 20	90 x 45	60 x 50	82 x 21	63 x 50	20 x 20
Maximum Resolution* (Å)	1.61	1.60	1.97	1.50	2.0	2.11
Total Angular Range (°)	130	105	240	90	180	70
Oscillation (°)	0.10	0.15	0.10	0.10	0.50	0.10
Exposure time per frame (s)	0.10	0.15	0.10	0.10	0.50	0.10
Transmission (%)	5	15	47	20	10	

6.2.4. Structure Solution and Refinement

Diffraction datasets were processed with the Xia2 3daii pipeline at Diamond with 5% of reflections removed for cross-validation (Kabsch, 2010b, Winter et al., 2013). All structures were solved by molecular replacement using either Molrep (Vagin and Teplyakov, 2010a) or MrBump (Keegan and Winn, 2008). The apo structure of *PA*3825-EAL was solved using residues 493-743 of a putative EAL domain homology model from *Thiobacillus denitrificans* (PDB code 2R6O) (Tchigvintsev et al., 2010). Metal-apo, metal-CdG and metal-pGpG complexes were solved using the refined apo structure as the search model. Further refinement of structures was performed with REFMAC5 (Murshudov et al., 2011), with model building carried out in COOT (Emsley et al., 2010). Model

validation was performed using Molprobity (Chen et al., 2010). Figures of all structures were prepared with PyMol version 1.3r1 (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC).

6.2.5. Size Exclusion Chromatography Coupled With Multi-angle Light Scattering

A SEC-MALS setup with a Superdex 200 10/300 GL SEC column on an AKTA Purifier (GE Healthcare) connected to DAWN HELEOS-II MALS and Optilab T-rEX refractive index detectors (Wyatt) was used to determine *PA*3825-EAL's oligomeric state under different conditions in solution. The column was equilibrated overnight in SEC buffer to give stable baseline signals from all detectors. Detectors were calibrated using Ribonuclease A according to standard Wyatt protocols.

SEC-MALS experiments consisted of 50 µl injections of either *PA*3825-EAL or MucR-EAL (Provided by Andy Hutchin, University of Sheffield) at varying concentrations to assess the effects of concentration on the oligomeric state. *PA*3825-EAL and MucR-EAL were also tested in the presence of excess substrate (5 mM CdG) and/or different divalent cations (2 mM $CaCl_2$, $MgCl_2$ or $MnCl_2$). All experiments were performed at 4°C at a flow rate of 0.3 ml/min. Substrate and divalent metals were added directly to the protein solution prior to loading samples onto the column. Molecular masses were calculated with a dn/dc value of 0.1825.

6.2.6. Enzymatic Assay

Enzymatic assays were carried out by assessing *PA*3825-EAL conversion of CdG to pGpG by ion exchange chromatography with a 1 ml Resource-Q column packed with 15 µm polyetheretherketone (PEEK) beads (GE-Healthcare). *PA*3825-EAL protein was stripped of all metals by adding 50 mM EDTA, which was later removed by dialysis to leave protein in 50 mM tris pH 9.3 and 50 mM NaCl. Enzymatic reactions were carried out in the presence of excess CdG (40 µM) and

various divalent metals (2 mM $CaCl_2$, $MgCl_2$, $MnCl_2$ or both $MgCl_2$ and $MnCl_2$), with reactions started by adding 1 µM *PA*3825-EAL to substrate at a final reaction volume of 600 µl. 100 µL fractions were removed after 0, 2.5, 5 and 20 minutes and reactions stopped by addition of 10 µl of 100 mM $CaCl_2$. Inhibited samples were made up to 1 ml with 890 µL of 5 mM ammonium bicarbonate (NH_4CO_3) and 100 µl loaded onto a 1 ml Resource-Q column pre-equilibrated in 5 mM ammonium bicarbonate. GdG and pGpG were eluted with a linear gradient of 5 mM to 1 M ammonium bicarbonate over 18 column volumes; elution was monitored by absorption at 254 nm. Optimum pH for PA3825-EAL in the presence of manganese was measured by the same protocol described above, with reactions between pH 4 and pH 8 stopped after 20 minutes.

Michaelis-Menten kinetics were assessed in the presence of manganese by assessing conversion rates of CdG to pGpG between 20 µM and 200 µM CdG by 0.2 µM *PA*3825-EAL over forty minutes. Reactions were initiated as detailed above with aliquots of all three CdG concentration fractions stopped at certain time points with $CaCl_2$. Samples were prepared and run down the Q-column as described above and the peak area for the pGpG peak calculated to be plotted against the Michaelis-Menten equation in order to determine the K_{cat}.

6.3. Results and Discussion

6.3.1. *PA*3825-EAL and MucR-EAL Enzymatic Activity

Both *PA*3825-EAL and MucR-EAL's ability to hydrolyse CdG to pGpG were assessed in the presence of magnesium, manganese and calcium. *PA*3825-EAL hydrolysed CdG in the presence of magnesium and manganese, with Mn^{2+}-loaded *PA*3825-EAL showing a much faster conversion rate compared to Mg^{2+} loaded. As expected, Ca^{2+}-loaded *PA*3825-EAL was completely inhibited. The catalytic rates of EAL domains has been found to be highly dependent on pH, with activity significantly increasing towards more alkaline pHs to a peak at around pH 9.3 (Schmidt et al., 2005). Analysis of the optimum pH for Mn^{2+}-loaded *PA*3825-EAL activity over 20 minutes shows conversion

rates are significantly reduced below neutral pH, with activity all but inhibited at pH 5 and below over a reasonable experimental timeframe (Fig. 6.14). *PA*3825-EAL shows a significant jump in activity from pH 6 (red) to pH 7 (light blue), with a further increase in activity in more alkaline conditions (orange). Measurements at pH 9 and above over the same time period showed no difference to measurements at pH 8.

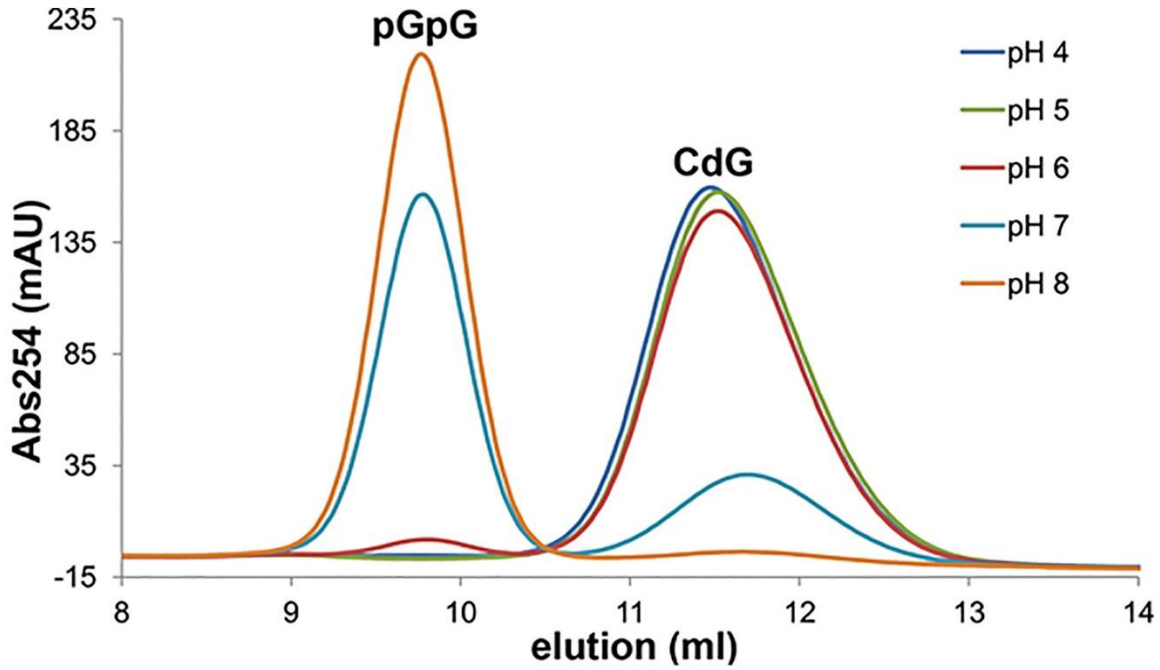


Fig. 6.14: FPCL chromatographs at 254 nm showing the pH dependent catalysis of PA3825-EAL. 40 uM CdG was added to 1 uM PA3825-EAL in the presence of 5 uM Mn^{2+} and reactions stopped with Ca^{2+} after 20 minutes.

*PA*3825-EAL at pH 9.35 in the presence of Mn^{2+} turns over substrate with a K_{cat} of ~ 0.85 s^{-1} (Fig. 6.15), a rate comparable with that of other EAL domains in the literature (Christen et al., 2005, Barends et al., 2009, Rao et al., 2008, Schmidt et al., 2005, Sundriyal et al., 2014). Mg^{2+}-loaded *PA*3825-EAL on the other hand showed a 30-40 times slower reaction rate compared to Mn^{2+}-loaded *PA*3825-EAL. Variation in the initial catalytic rate of *PA*3825-EAL at different substrate concentrations could not be observed, as seen in figure 6.15b, even at the lowest substrate concentration detectable by the experimental setup (5 µM) (Fig. 6.16). Based on these observations the K_m value must lie below 4 µM, in agreement with other recently published EAL domain activity (Sundriyal et al., 2014).

Unlike *PA*3825-EAL, MucR-EAL showed no activity in the presence of magnesium, manganese of calcium, in agreement with the observation that his-tag purified MucR-EAL contained high amounts of bound nucleotide, as judged by absorption at 270 nm (private communication from the Tews group, University of Southampton). Bound nucleotide was later revealed as CdG by the high resolution crystal structures. Based on sequence alone the PDE inactivity of MucR in solution is surprising as all necessary catalytic residues are conserved, as seen in other PDE active EAL domains (Fig. 6.17). The only significant difference observed between *PA*3825-EAL and MucR-EAL that may account for MucR's inactivity, is that while *PA*3825-EAL is almost completely monomeric in solution while MucR is exclusively dimeric.

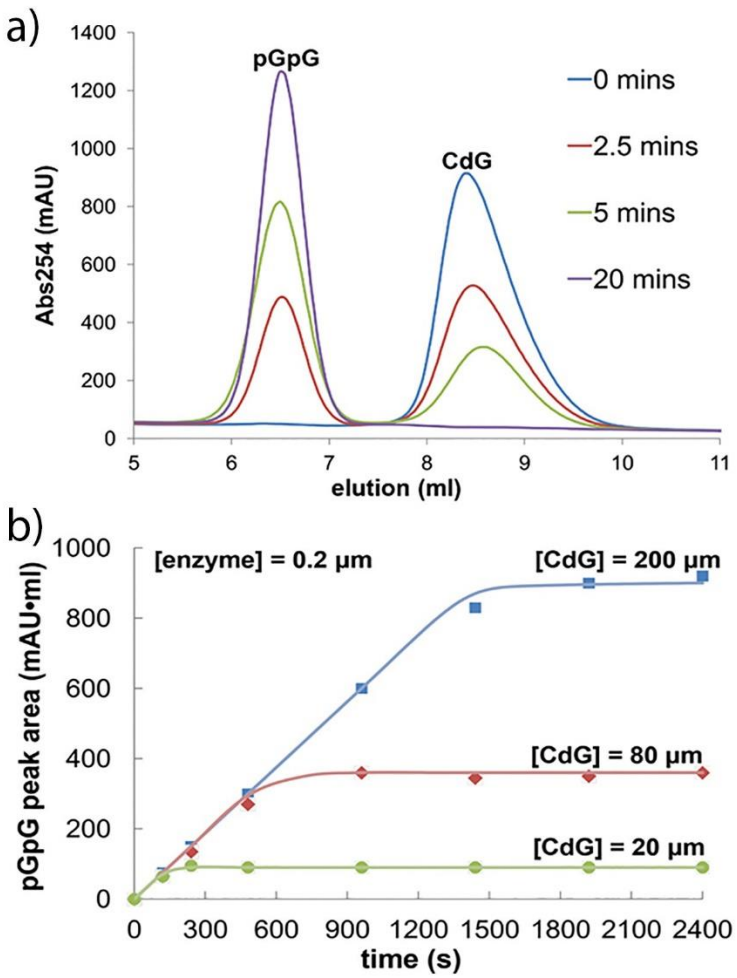


Fig. 6.15: *PA*3825-EAL catalytic rates. a) FPLC trace showing time dependent conversion of 500 uM CdG to pGpG by 5 uM *PA*3825-EAL in the presence of Mn^{2+}. b) Progress curves showing conversion of CdG to pGpG at various concentrations as indicated by 0.2 uM PA3825-EAL. pGpG peak area data were fitted to a simple Michaelis-Menten kinetics model.

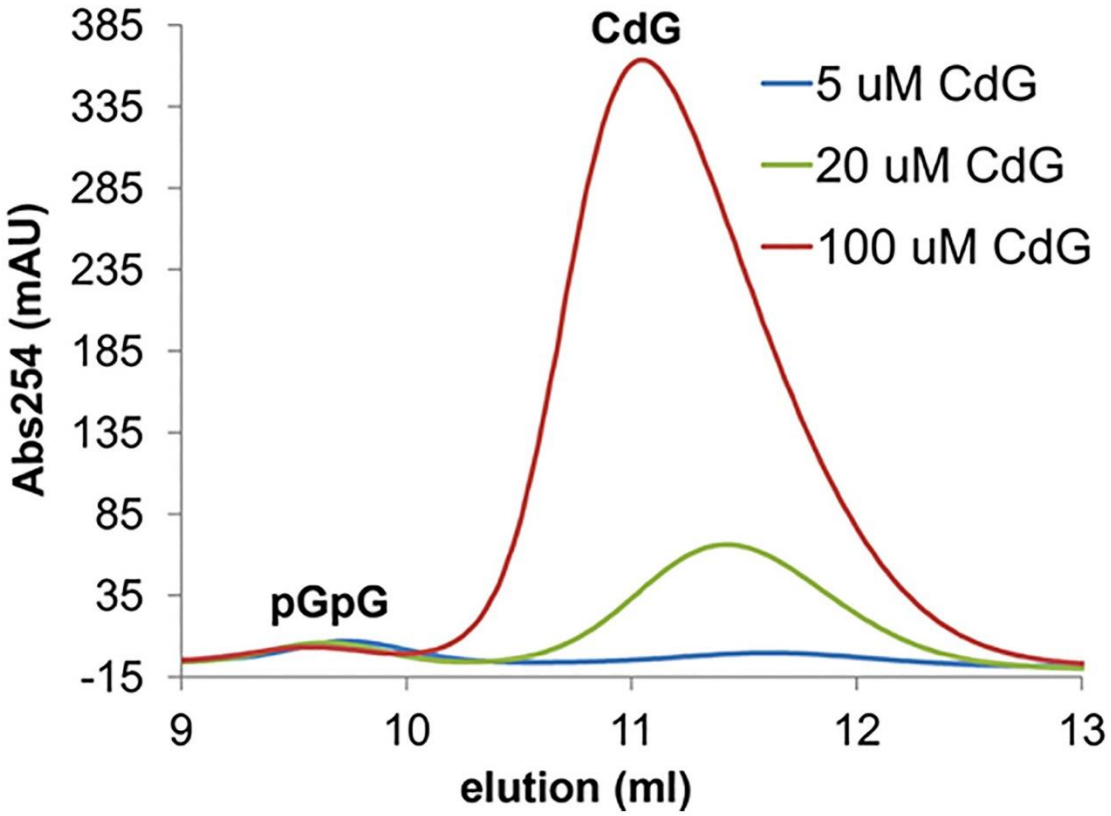


Fig. 6.16: The Minimum amount of CdG detectable through FPLC methods to determine initial catalytic rates of *PA*3825-EAL hydrolysis of CdG to pGpG. 0.08 uM *PA*3825-EAL in the presence of Mn^{2+} was incubated with specified CdG concentrations for one minute before running through a resource Q column.

6.3.2. *PA*3825-EAL and MucR-EAL Dimerisation and Activity

The oligomeric states of various EAL domains in crystal structures and/or in solution have been variously reported as monomers (Schmidt et al., 2005), dimers (Barends et al., 2009, Tchigvintsev et al., 2010, Minasov et al., 2009, Rao et al., 2008) and tetramers (Rao et al., 2008, Chen et al., 2012). SEC-MALS data were collected to establish the oligomeric states of *PA*3825-EAL and MucR in solution. *PA*3825-EAL samples at 3 mg/ml and 6 mg/ml show a mixture of monomers and dimers with monomer becoming even more prevalent with increasing concentration (Fig. 6.18a). This finding is counter intuitive and may warrant further investigation. Addition of both activating and inhibiting metals (Mg^{2+} and Ca^{2+}), with or without CdG, saw a slight shift towards the monomeric state (Fig. 6.18b). Both conditions lead to formation of the catalytic ternary complex, suggesting that catalytic reactions in *PA*3825-EAL favour the formation of monomers. Despite this prevalence as a

monomer in solution, analysis of macromolecular interfaces using the PISA server (Krissinel and Henrick, 2007) suggest a dimer as the biological assembly for the tetragonal form of *PA*3825-EAL, while the monoclinic crystals showed no stable quaternary complex structures were likely. PISA's putative though unstable interface predicted for the monoclinic form, varies considerably from the predicted stable interface for the tetragonal crystal (Fig. 6.19). Figure 6.19 shows the different predicted interfaces of the tetragonal (Fig. 6.19a) and monoclinic (Fig. 6.19b) crystal structures. These differences, along with the SEC-MALS data suggest *PA*3825-EAL does not form tight interacting dimers, strongly suggesting that stable tetragonal dimer is most likely a product of the crystal packing arrangement as opposed to a functional dimerisation.

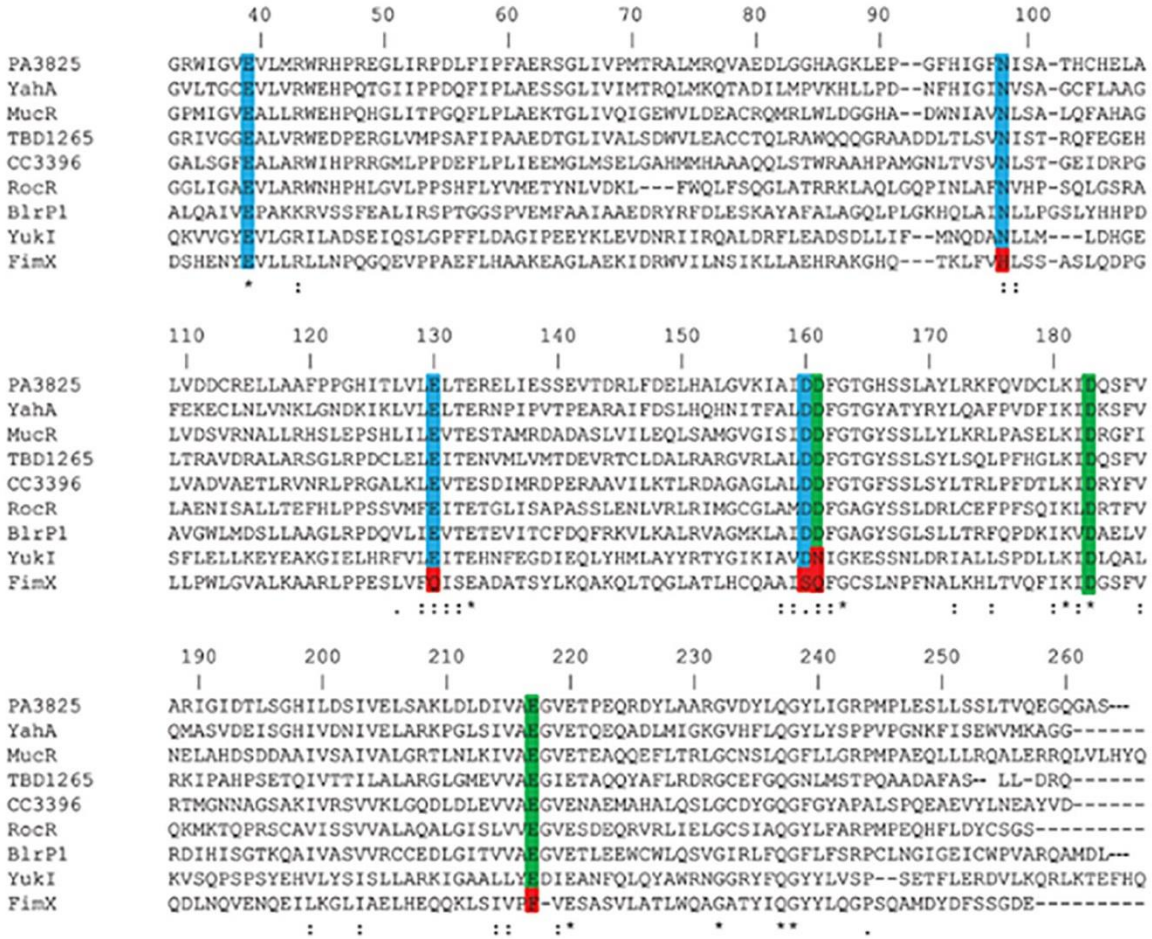


Fig. 6.17: Multiple sequence alignment of *PA*3825-EAL to other well characterised active and inactive EAL domains. M1 and M2 coordinating residues are coloured in cyan an green, respectively., and variations in the conserved motifs are highlighted in red.

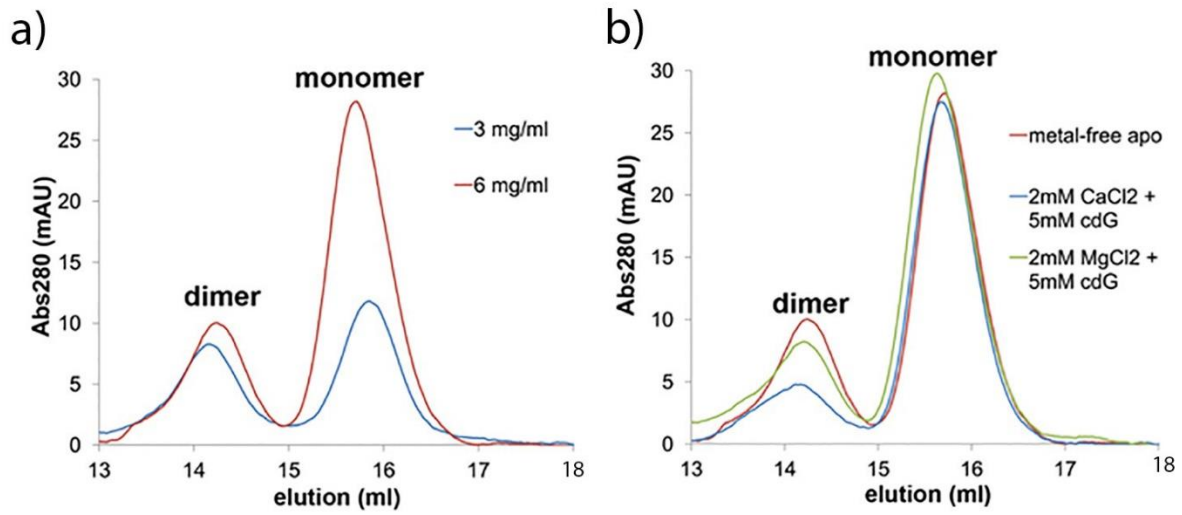


Fig. 6.18: FPLC chromatogram of *PA*3825-EAL oligomeric state analysis with SEC-MALS. a) Concentration dependent oligomerisation gel filtration elution peaks with molecular weights of 30 and 60 kDa, respectively for monomeric and dimeric states, as estimated by MALS. b) Gel filtration peaks of 6 mg/ml *PA*3825-EAL as a metal-free apo and in the presence of CdG with inhibiting calcium and activating magnesium as indicated.

In stark contrast to *PA*3825-EAL, MucR elutes from the gel filtration column as a single peak with a molecular weight comparable to dimeric MucR, as estimated by MALS. PISA calculations gave a stable quaternary structure with a dissociation constant of 7.0 kcal/mol, similar to the 11.2 kcal/mol predicted for tetragonal *PA*3825-EAL. However, the solvation free energy on formation of MucR is five times more negative than that of *PA*3825-EAL and the MucR dimer has twice the buried surface area. The dimerisation interfaces between MucR monomers and *PA*3825-EAL monomers are completely different from one another in sequence (Fig. 6.19c). Based on these observations and SEC-MALS data it is reasonable to conclude that the dimerisation domain of MucR is real and not an artefact of crystal packing. Despite this, MucR is still found to be inactive in solution and with CdG in the active site of the crystal structures. Based on these observations there is evidence to suggest a general role of dimerisation inhibition in EAL domains. Further evidence of dimerisation based inhibition of EAL domains is presented in later sections with the structure of MucR-EAL.

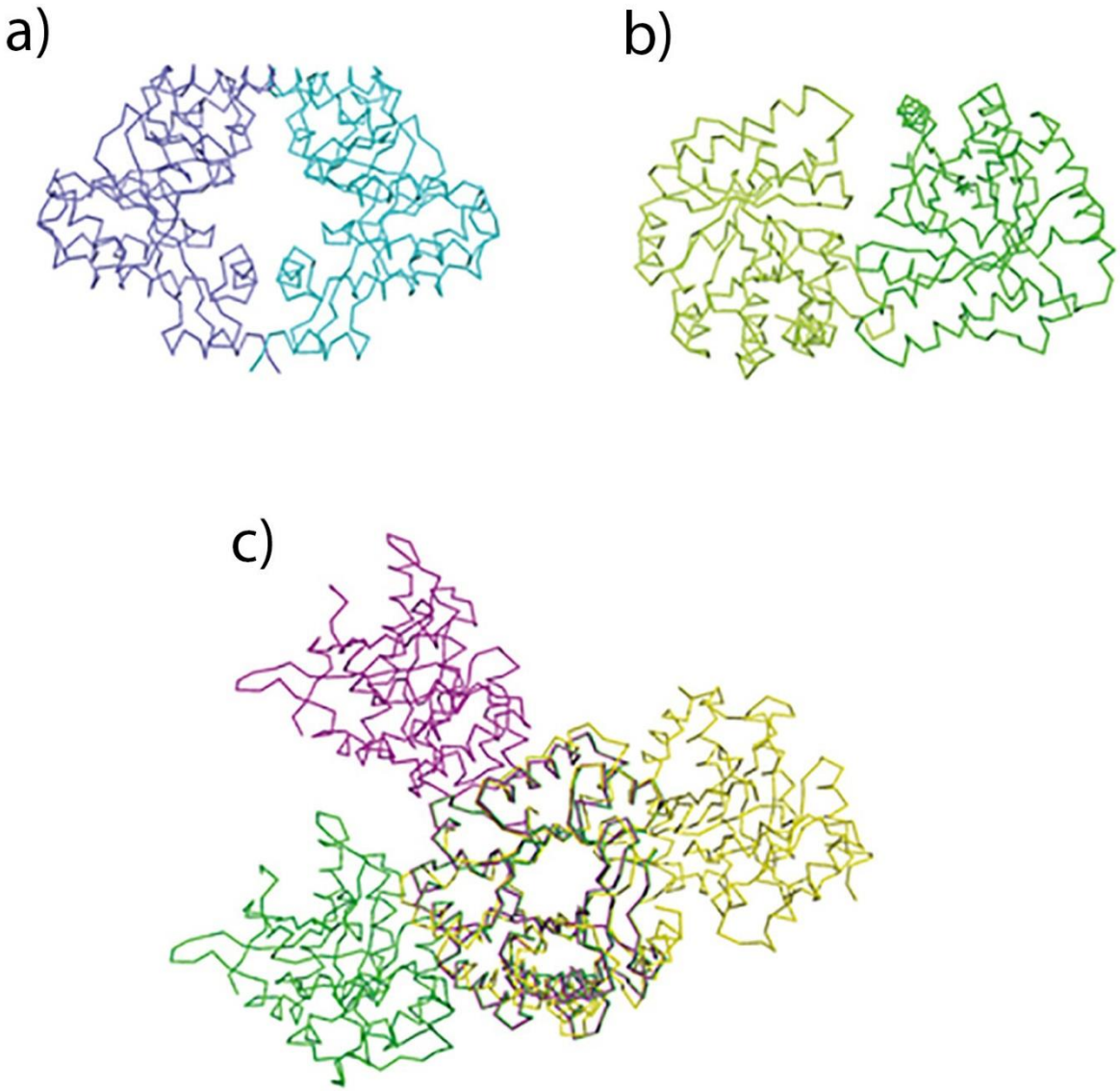


Fig. 6.19: DImerisation interfaces of EAL domain crystal structures. All structures are shown as ribbons. a) Tetragonal *PA*3825-EAL dimer generated through crystal symmetry (4Y9M, 4Y9O and 4Y9Q). b) Monoclinic *PA*3825-EAL dimer (4Y8E and 4Y9P). c) superposition of chain A dimers of *PA*3825-EAL in monoclinic crystals (yellow), tetragonal crystals (green) and MucR-EAL (magenta).

6.3.3. *PA*3825-EAL Apo and CdG-bound Structures

High resolution structures of metal-free-apo, calcium apo, calcium CdG, magnesium CdG and manganese CdG *PA*3825-EAL complex were obtained; along with pGpG bound *PA*3825-EAL which will be presented separately in the next section. Data processing, refinement and validation statistics for all structures are presented in Table 6.2. The triosephosphate isomerase (TIM)-barrel

fold (αβ(βα)7) of *PA*3825-EAL is conserved, along with the binding of CdG in the typical extended conformation at the C-terminal end of the β-barrel (Fig. 6.20a and 20b, respectively). Moreover, *PA*3825-EAL structures and substrate binding are consistent with previously published CdG-EAL domain complexes, including BlrP1 (Barends et al., 2009), YahA (Sundriyal et al., 2014), FimX (Robert-Paganin et al., 2012) and RocR (Chen et al., 2012). The overall fold of *PA*3825-EAL remains unchanged between structures with metal ions and no substrate and structures with metal ions and CdG. The only observable differences occur between metal-free-apo structures and metal-apo/metal-CdG structures. β-strands 1 and 2 (residues 22-30 and 37-44, respectively) are presented as disordered loops in metal-free-apo structures, becoming ordered in the presence of metal ions to complete the TIM-barrel-fold (Fig. 6.21). Without metal and/or substrate bound Leu27, Glu39, Leu41 and Arg43 all present with double conformations (green in Fig. 6.21), which stabilise upon metal binding (purple in Fig. 6.21). The second difference arises in loop 6 (residues 160-170) which shows a significant shift in position upon metal binding, moving into the active site to coordinate the catalytic metal ions with two highly conserved Asp residues (Asp160-161) (Fig. 6.22).

Mg^{2+} and Ca^{2+} CdG-bound *PA*3825-EAL (Fig. 6.23a and 6.23b respectively) present with two metals in the catalytic centre, labelled M1 and M2, as observed in similar EAL domain structures in the PDB (Fig. 6.24). Although Mg^{2+} is an activating metal in EAL domains, Mg^{2+}-loaded *PA*3825-EAL PDB (Fig. 6.24). Although Mg^{2+} is an activating metal in EAL domains, Mg^{2+}-loaded *PA*3825-EAL co-crystallisation with CdG presented with an intact CdG molecule in the binding pocket (Fig. 6.23a); which is unsurprising considering Mg^{2+}-loaded *PA*3825-EAL is highly inefficient catalytically at the

Table 6.2: Crystallization, crystallographic data collection and structure refinement statistics. Values in parentheses refer to the highest resolution bin.

Structure	**PA3825-EAL Apo-Apo**	**PA3825-EAL Ca-Apo**	**PA3825-EAL Ca-CdG**	**PA3825-EAL Mg-CdG**	**PA3825-EAL-Mn-CdG**	**PA3825-EAL Na-pGpG**
Data Collection						
Space Group	$P4_32_12$	C2	C2	$P4_32_12$	$P6_1$	$P4_32_12$
Cell Axes (Å)	a = 64.34, b = 64.34, c = 134.00	a = 112.4, b = 59.8, c = 96.7	a = 112.2, b = 59.4, c = 92.8	a = 64.6, b = 64.6, c = 135.7	a = 107.9, b = 107.9, c = 93.5	a = 64.8, b = 64.8, c = 135.8
Angles (°)	α = β = γ = 90	β = 115.0	β = 115.0	α = β = λ = 90	γ = 120	α = β = γ = 90
Beamline	I03	I04	I04-1	I02	I04	I02
Wavelength (Å)	0.9763	0.9795	0.9200	0.9795	0.9795	1.8785
Resolution (Å)	64.34-1.60 (1.64-1.60)	43.10-1.61 (1.65-1.61)	51.29-2.44 (2.50-2.44)	29.16-1.77 (1.81-1.77)	33.06-2.59 (2.66-2.59)	27.16-2.15 (2.21-2.15)
Unique Reflections (#)	36008 (2598)	71975 (5229)	20405 (1494)	28763 (2072)	26152 (1919)	16288 (1689)
Redundancy	9.3 (9.2)	3.0 (3.1)	4.5 (4.8)	6.2 (6.7)	10.3 (10.5)	3.8 (3.8)
R_{pim} (%)	3.1 (30.2)	2.4 (38.9)	2.6 (42.2)	2.2 (26.7)	5.4 (94.2)	7.2 (40.5)
I/σ(I)	18.0 (3.2)	18.3 (2.3)	17.3 (2.0)	18.8 (2.6)	17.5 (1.3)	6.6 (2.4)
Completeness (%)	100.0 (99.5)	97.4 (96.4)	98.1 (98.9)	99.5 (99.8)	99.9 (99.7)	98.8 (99.2)
CC1/2	0.999 (0.900)	0.999 (0.761)	0.999 (0.774)	0.999 (0.851)	0.999 (0.756)	0.988 (0.725)
Refinement						
Molecules/AU	1	2	2	1	2	1
R_{work}/R_{free} (%)	18.3/23.2	18.7/22.3	21.3/27.2	19.1/23.7	18.7/25.3	20.0/26.1
Rmsd						
Bond Length (Å)	0.022	0.022	0.012	0.020	0.012	0.020
Bond Angles (°)	2.088	2.128	1.717	2.002	1.684	2.343
B Factors ($Å^2$)						
Average B factor	22	33	67	33	49.5	36
Ramachandran Statistics (%)						
Favoured Regions	99.3	98.2	93.8	98.2	95.5	97.3
Allowed Regions	0.7	1.8	5.2	1.4	4.3	2.7

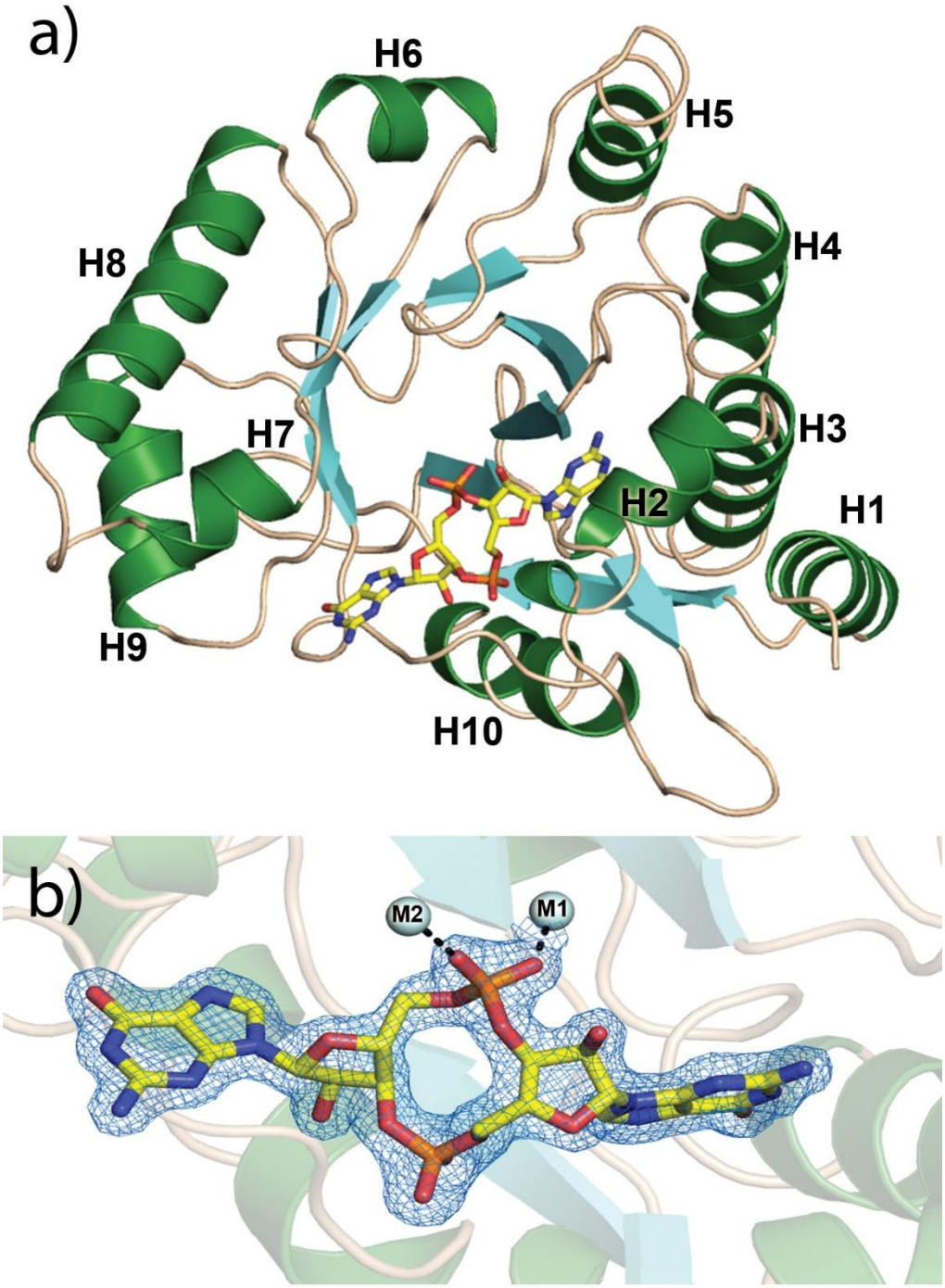


Fig. 6.20: Structure of *PA*3825-EAL. a) Structure of the magnesium-loaded PA3825-EAL monomer with α-helicies in green and β-sheets in cyan. The bound substrate CdG is shown as yellow sticks. b) A detailed view of the bound substrate with $2F_o$-F_c density contoured to 1.8 σ. The two catalytic metals, in this case Mg^{2+}, are labelled M1 and M2.

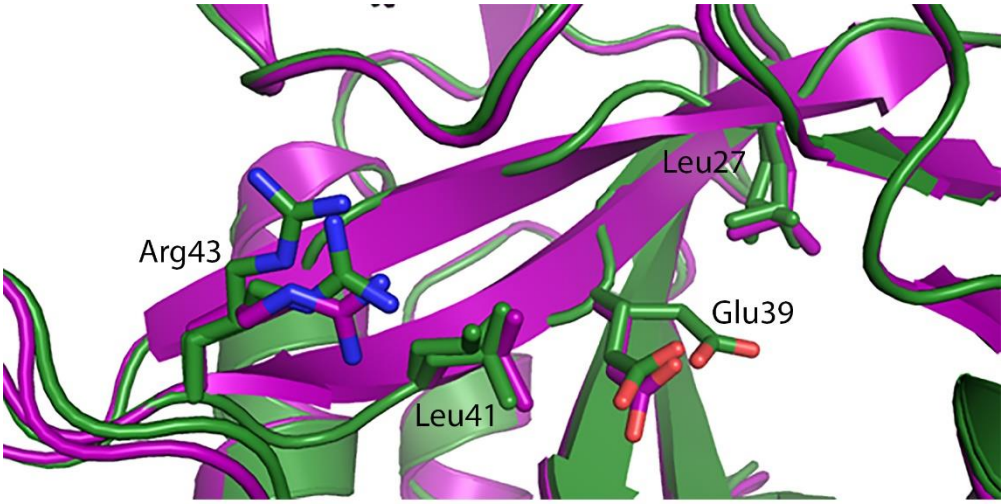


Fig. 6.21: Detailed view of PA3825-EAL changes in secondary structure with and without metal ligands. Metal-free apo PA3825-EAL (green) presents with double conformations causing ordered β-sheets observed in metal-apo and metal-ligand (purple) structures to become disordered loops. Both structures are presented as cartoons with double conformation and corresponding single conformation residues presented as sticks coloured by element.

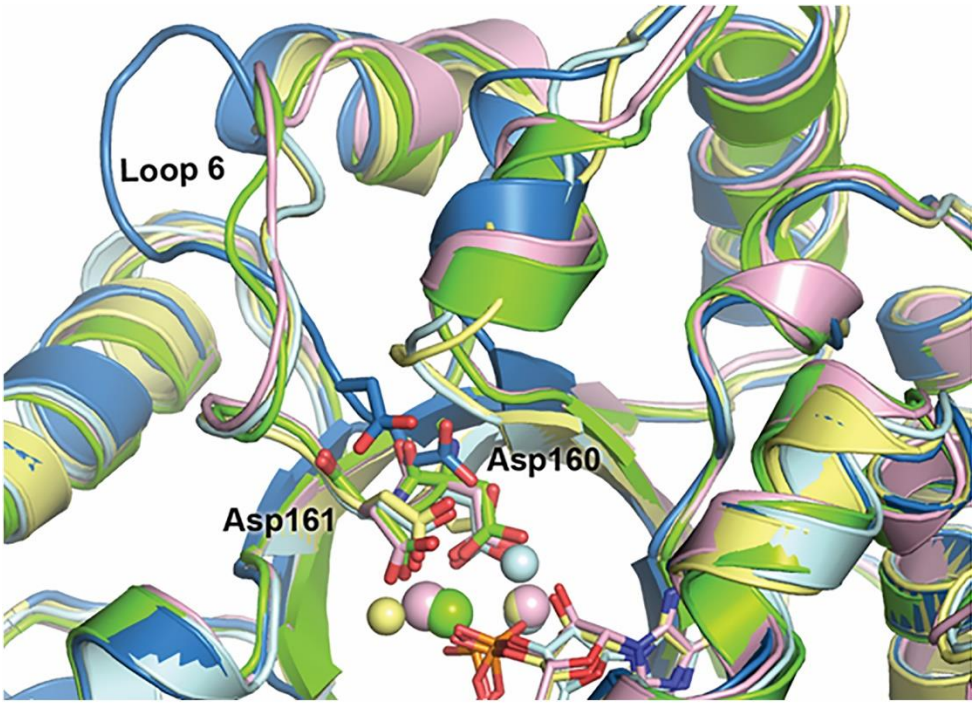


Fig. 6.22: Superposition of metal-free apo, metal-apo, metal-substrate and metal-product bound PA3825-EAL structures. Metal-free apo (blue), calcium-apo (green), calcium-CdG (pink), magnesium-CdG (yellow) and manganese-pGpG (pale cyan) PA3825-EAL structures are represented as cartoons. Catalytic Asp residues (160-161), CdG/pGpG and metal ions are represented as sticks, lines and spheres, respectively. Residues and ligands are coloured by element and metal ions to corresponding structures. The catalytic loop 6 is indicated.

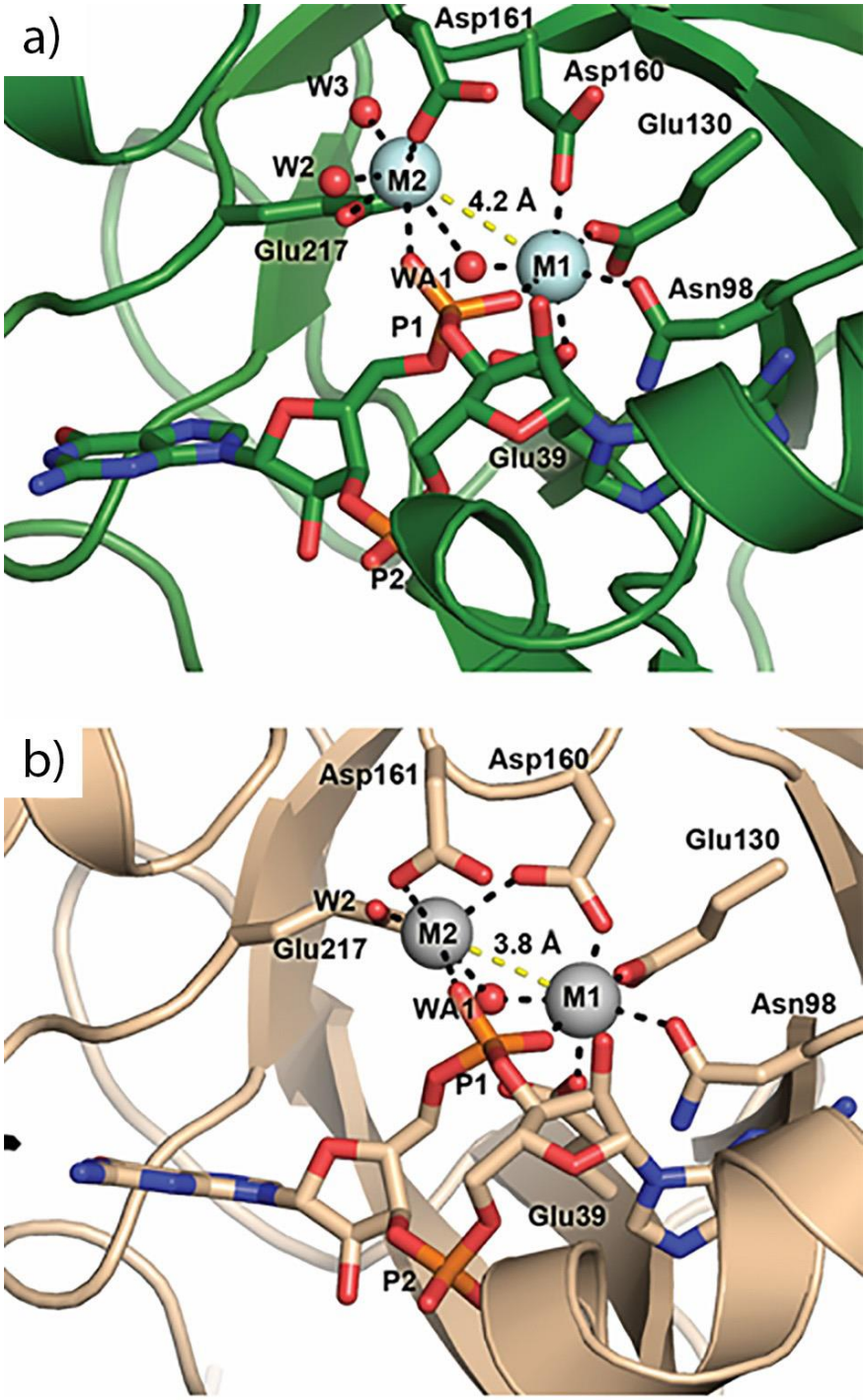


Fig. 6.23: A detailed view of the coordination spheres of the bimetallic centre of magnesium- (a) and calcium- (b) loaded *PA*3825-EAL structures. Protein-metal interactions and inter-metal distances are shown as black and yellow dashed lines, respectively. Protein-metal liganding residues and CdG are shown as sticks and coloured by atom type. Ca^{2+}, Mg^{2+} and water molecules are shown as pale cyan, grey and red spheres respectively.

crystallisation pH (pH 7.1). Although metal site M1 shows 100% occupancy, the more mobile M2 site in both Ca^{2+} and Mg^{2+} show a lower occupancy, of approximately 30%. The metal ion occupying the

M1 site, either Mg^{2+} or Ca^{2+}, is octahedrally coordinated by Glu39, Asn98, Asp130, Asp160, a free oxygen from the scissile phosphate (P1) of CdG and the putative in-line attacking water molecule, WA1 (Fig. 6.23a and 6.23b). In the case of Mg^{2+}-loaded *PA*3825 EAL, the M2 magnesium ion is octahedrally coordinated by Asp161, Glu217, a P1 free oxygen and three waters, WA1, W2 and W3 (Fig. 6.23a). On the other hand, in Ca^{2+}-loaded *PA*3825 EAL the M2 calcium ion is trigonal bipyramidally coordinated by Asp160, Asp161, Glu217, a P1 free oxygen and one water, W2 (Fig. 6.23b). It should be noted that in the Mg-CdG structure Asp160 coordinates to metal M1 only, rather than bridging both metal sites through the carboxylate fork as characteristically observed in bimetallic EAL domains (as well as in Ca-CdG *PA*3825-EAL). This variability in coordination may explain the poor occupancy and larger inter-metal distance of the M2 site. The position of the M1 site is perfectly matched between Mg^{2+}-*PA*3825-EAL and Ca^{2+}-*PA*3825-EAL, as observed by superposition of the two structures; whereas the M2 site is displaced by approximately 0.4 Å between Mg^{2+}-*PA*3825-EAL and Ca^{2+}-*PA*3825-EAL (Fig. 6.23a and 6.23b). A review of other metal-bound structures of EAL domains in the PDB shows there is a significant variation in the inter-metal M1-M2 distances (Fig. 6.24). The largest inter-metal distance observed between bimetallic substrate bound structures is 6.0 Å in Mg-*PA*3825-EAL (Fig. 6.24a) while the smallest distance, 3.9 Å, occurs in Mn-BlrP1 (Fig. 6.24n). All available bimetallic EAL domain structures, as of April 2015, are presented in figure 6.24. Bimetallic EAL structures have been solved over a range of activating and inhibitory pHs (4.5-9.0) and in the presence of activating and inhibiting metal ions. Overall there appears to be little correlation between EALs being active (Mg^{2+}/Mn^{2+} ≥ pH 7.0) or inactive (Mg^{2+}/Mn^{2+} < pH 7.0/ Ca^{2+} at any pH) and inter-metal distance. Inter-metal distances for Ca^{2+} inhibited structures vary considerably from 5.9 Å to 3.9 Å, although active Mn^{2+} structures are more limited in their variability, between 4.5 Å and 3.9 Å.

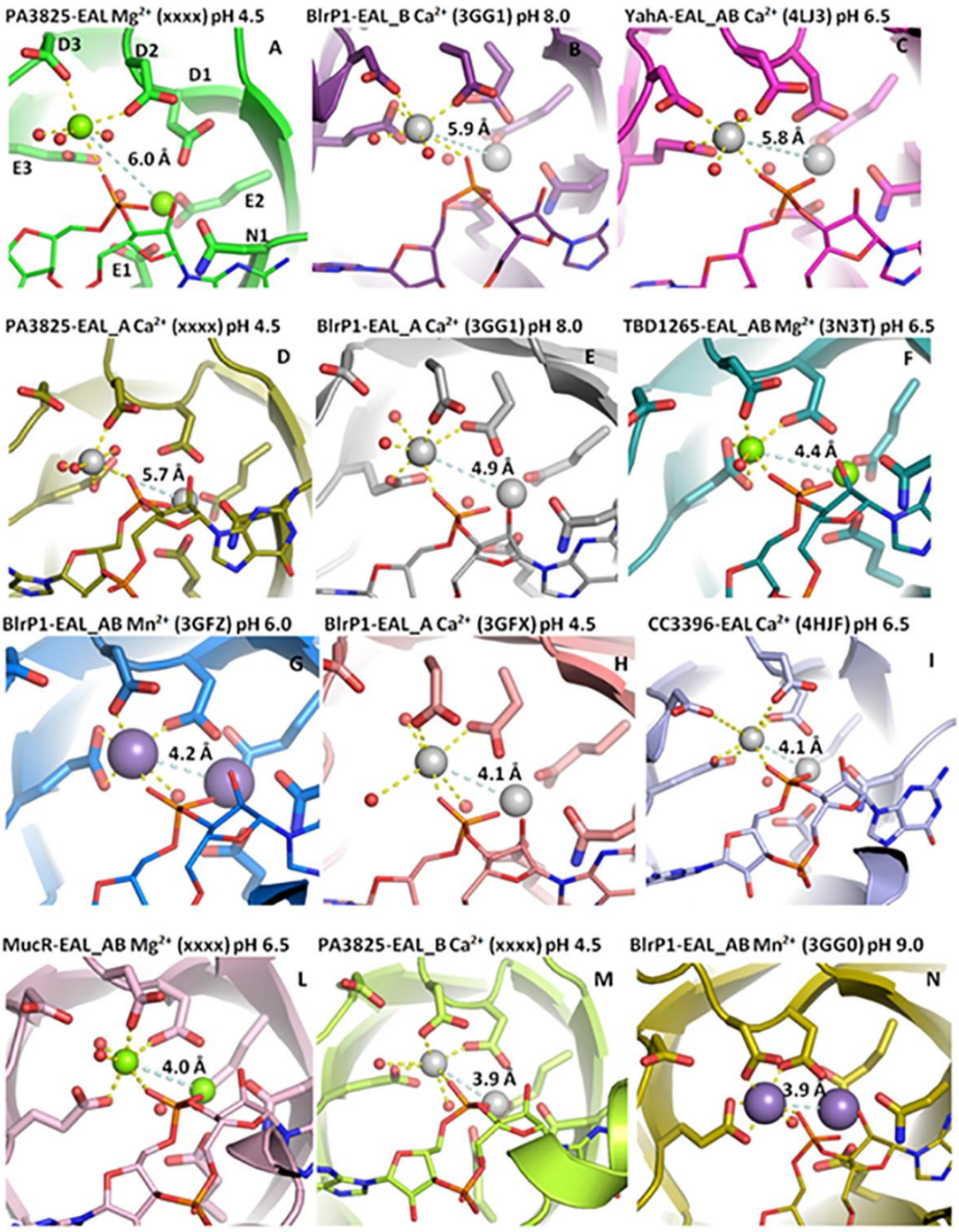


Fig. 6.24: Inter-metal distances of all bimetallic EAL domains in the PDB arranged by decreasing inter-metal distance and labelled with protein name, PDB code and crystallisation pH. Conserved residues are labelled in (a) only for clarity. Metal liganding residues are shown as sticks and ligands as lines with both coloured by atom type and corresponding structure. Calcium, magnesium and manganese ions are presented as grey, green and purple spheres, respectively. Protein-metal distances an inter-metal distances are represented by yellow and blue dashed lines, respectively. Dimeric structures with the same inter-metal distances between molecules are labelled as AB and structures with differing distances between molecules are labelled A and B accordingly.

Two previously published structures of Mn-CdG-Blrp1 (table 2) were crystallised at pH 6 and pH 9 (PDB codes 3GFZ and 3GG0, respectively), representing inactive and fast acting enzymes based on *in vitro* kinetic analysis of Blrp1 at corresponding pH values (Barends et al., 2009). These structures saw the inter-metal distance decrease with increasing pH from 4.2 Å to 3.9 Å, a difference of 0.3 Å. Based on this, the authors proposed a pH-dependent catalytic regulation mechanism based on a "tight" metal centre whereby deprotonation of the in-line attacking water for nucleophilic attack only occurs when inter-metal distances reach a critical proximity. However, this theory suffers from a small sample size, consisting of a single Blrp1 structure at each pH. Structures were also determined to medium resolution (2.05 Å and 2.55 Å for pH 6 and pH 9 structures respectively), making it difficult to make arguments based on something as precise as a 0.3 Å difference in metal distance. The intrinsic variability of M2's position, even between chains of dimeric structures (Fig. 6.24b and 6.24e, 6.24d and 6.24m (*PA*3825 Mn structure), also brings such a mechanism into doubt. In addition to this, the structure of Mn-Cdg-*PA*3825-EAL at pH 7.1 shows a larger inter-metal distance than Mn-CdG-Blrp1 despite being more catalytically active. Finally, the fast acting Blrp1 structure crystallised in the presence of manganese at pH 9 displaying this "tight" conformation (PDB code 3GG0), shows an intact CdG molecule after being soaked in CdG for 24 hours. *In vitro* analysis of Blrp1 in solution in the presence of manganese at pH 9 shows conversion of CdG to pGpG with a k_{cat} of approximately 7 s^{-1} (Barends et al., 2009), ample time for conversion of all available CdG to pGpG over 24 hours. A novel mechanism involving a new third metal site in the catalytic pocket, not observed in Blrp1 structures, may account for this paradox.

6.3.4. *PA*3825-EAL pGpG Structure

Soaking and co-crystallisation of tetragonal *PA*3825-EAL crystals with CdG and $MnCl_2$ gave significantly different results. Crystal structures following soaking with 10 mM CdG and 0.2 M $MnCl_2$

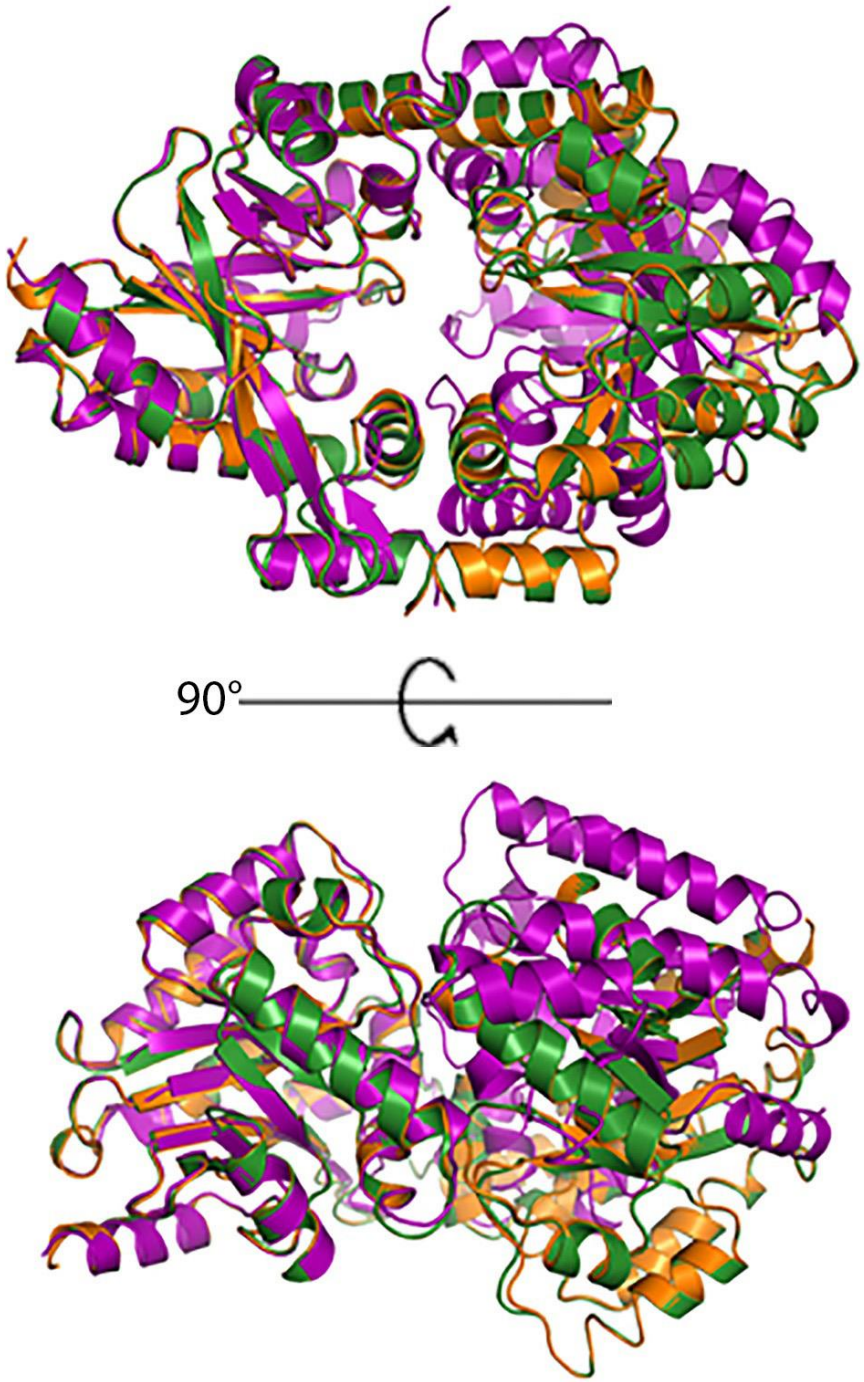


Fig. 6.25: Differences in dimerisation interfaces between long soak hexagonal Mn-CdG *PA*3825-EAL crystals (purple), tetragonal Mn-CdG *PA*3825-EAL co-crystallisation crystals (green) and tetragonal Mg-CdG *PA*3825-EAL co-crystallisation crystals (orange). Crystal structures are superimposed against A chains, with all structures presented as cartoons.

over ten days resulted in CdG bound in the catalytic pocket, with a fully occupied M1 site confirmed by anomalous manganese signal and the M2 site 20% occupied with no anomalous signal. Crystals

also underwent a space group change from tetragonal to hexagonal (Table 1) and a significant change in the dimeric interface *in crystallo* (Fig. 6.25), with hexagonal crystal B molecules (Fig. 6.25b,

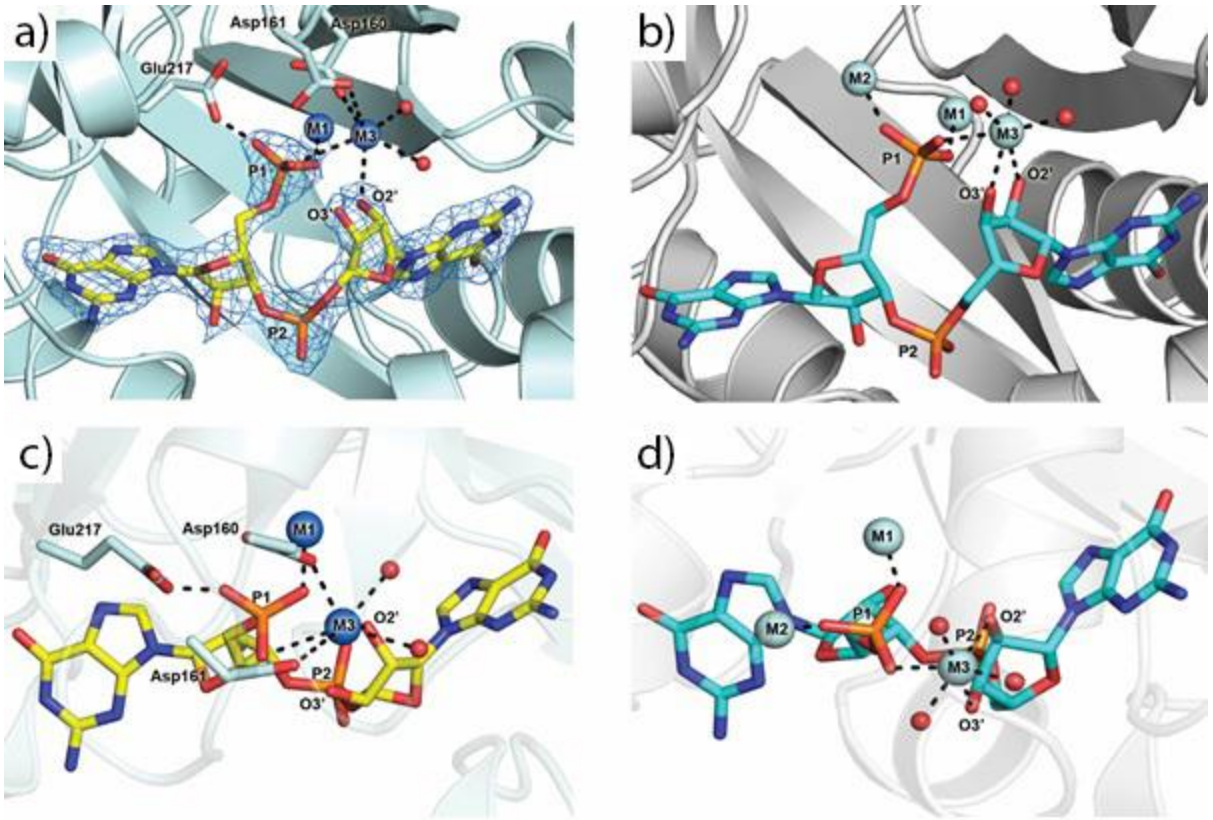


Fig. 6.26: Comparison of the novel M3 metal site observed in *PA*3825-EAL and *CC*3396-EAL in the presence of pGpG. a) Detailed view of *PA*3825-EAL's pGpG bound catalytic centre with $2F_o$-F_c electron density contoured to 1.4 σ at 2.27 Å resolution. Metal sites M1 and M2 have been modelled as low occupancy manganese ions. b) Detailed view of *CC*3396-EAL's catalytic centre with pGpG bound with three magnesium ions. c) and d) Overhead views of a) and b), respectively. Products and liganding residues are presented as sticks coloured by atom type, manganese and magnesium as dark blue and cyan spheres, respectively and all metal/ligand-pGpG interactions as black dashed lines.

purple) shifting by approximately 45° relative to the A molecule when compared to tetragonal structures (Fig. 6.25b, green and orange). Crystals soaked over hours or 1-7 days showed the initial tetragonal space group and no CdG In the active site. On the other hand, co-crystallisation of metal-free *PA*3825-EAL with CdG and Mn^{2+} over 4-5 weeks at 4°C yielded high resolution diffracting crystals, revealing that the phosphodiester bond of CdG coordinated by the catalytic metals had been cleaved, converting CdG (Fig. 6.22) to pGpG (Fig. 6.26a). This result is consistent with *in vitro* kinetics of Mn^{2+}-loaded *PA*3825-EAL at pH 7.1 (crystallisation pH), which shows significant catalytic

activity (Fig. 6.14). The reasons why soaking and co-crystallisation resulted in substrate and product bound structures respectively are unclear, although the lower temperature could significantly lower the activity of *PA*3825-EAL making 10 days too short a time span to see conversion of enough CdG *in crystallo* to give well defined electron density. Alternatively, cleavage of CdG may have occurred prior to crystallisation, when metal-free *PA*3825-EAL was mixed with 10 mM CdG and Mn^{2+}. Enzymatic activity may also have been inhibited by pseudo-dimerisation caused by crystal packing; a putative inhibitory dimerisation interface is discussed in more detail in the following section.

Cleavage of the scissile phosphoester bond in CdG caused significant changes in the ternary complex of *PA*3825-EAL, with the M2 site along with the coordinating waters disappearing completely and a new metal site (M3) appearing (Fig. 6.26a and 6.26c). The M1 site remains unchanged, with coordination exactly the same as that of calcium- or magnesium-CdG ternary complexes (Fig. 6.23a and 6.23b, respectively). Despite the presence of excess manganese in the crystallisation drop, both metal sites show a low occupancy of only 30%. This may stem from increased flexibility of the phosphate group following cleavage destabilising the ternary complex to allow release of the product. M3 also shows an octahedral coordination sphere consisting of Asp160, Asp161, a P1 free oxygen, the hydrolysed O2' Oxygen from the liganding sugar ring and two water molecules, WC4 and WC5 (Fig. 6.26a and 6.26c). Coordination of M3 to the P1 free oxygen and O2' oxygen of the sugar moiety is of particular interest as both oxygens are involved in the formation of the scissile phosphoester bond cleaved by *PA*3825-EAL.

The PDB has two more EAL domains reported in complex with pGpG, FimX-EAL from *P. aeruginosa* (4AFY) (Robert-Paganin et al., 2012) and CC3396-EAL from *C. crescentus* (3U2E) (unpublished). Inspection of the electron density maps of 4AFY based on structure factors shows considerable negative electron density for the pGpG bound to FimX-EAL (Fig. 6.27). In addition to extensive negative density, FimX-EAL is a degenerated EAL domain with a kcat of only 10^{-4} s^{-1} (Kazmierczak et al., 2006).

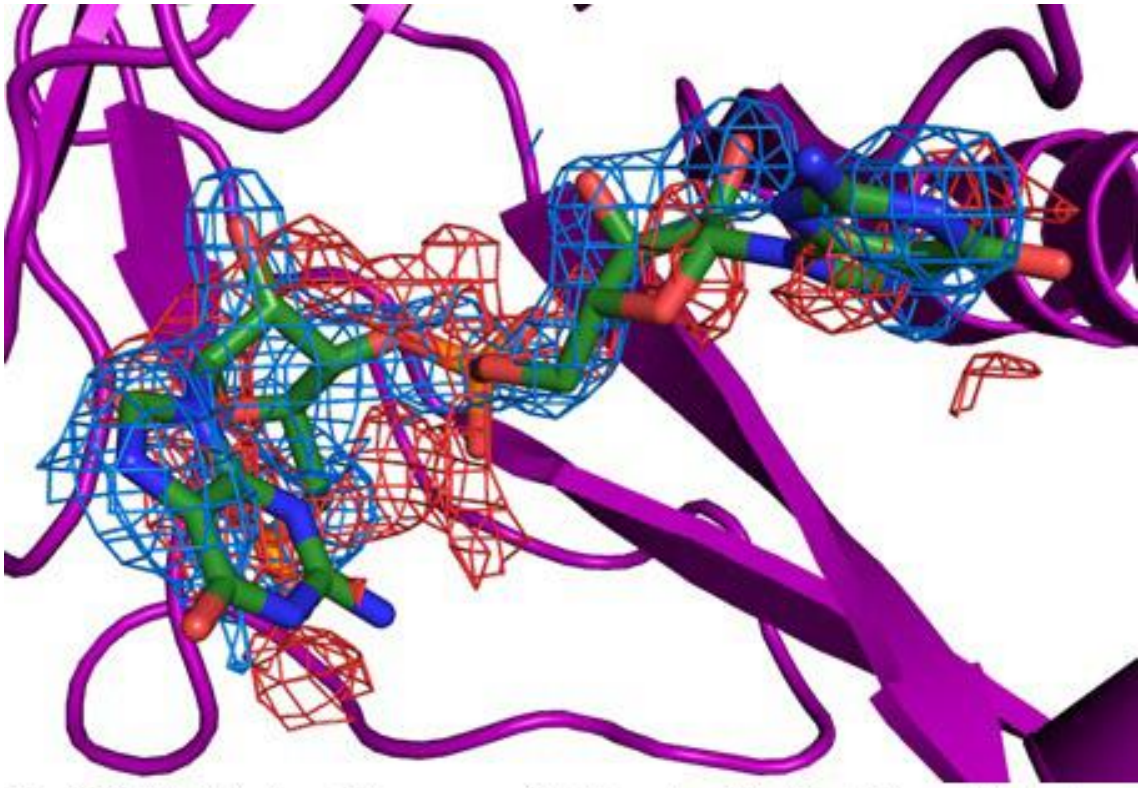


Fig. 6.27: Detailed view of the proposed binding site of the Fimx-EAL crystal structure (4AFY) with a putative pGpG molecule represented as green sticks and coloured by element. Blue density shows a $2F_o$-F_c map contoured to 1 σ and red density shows a negative F_o-F_c map contoured to 3 σ at 2.01 Å resolution.

The metal binding of all forty deposited EAL domain structures, both active and inactive and in the presence or absence of substrate, are presented in table 6.3. It is clear from table 6.3 that degenerated EAL domains such as FimX-EAL, LapD-EAL and YdiV have never been observed with bound metal in the active site, as expected from the mutations in the metal coordinating Asp residues (Fig. 6.17). Although catalysis is inhibited by these mutations the ability to bind CdG is retained, likely to allow the EAL domain to act as a sensory domain (Schmidt et al., 2005). It also appears that substrate is necessary to allow binding of M2 to the active site, which is never observed in metal-apo complexes. The conformation of pGpG in *PA*3825-EAL and *CC*3396-EAL are highly superposable (Fig. 6.28a), whereas 4AFY's pGpG varies greatly and presents with stereochemical clashing issues (Fig. 6.28c). As a result the structure of 4AFY has been disregarded in comparisons of pGpG binding in *PA*3825-EAL and *CC*3396-EAL.

Table 6.3. Classification of all EAL domain structures in the PDB to date. Inactive EAL domains are indicated in italics.

no metal CdG	no metal no substrate	1 metal (M1) no substrate	1 metal (M1) CdG	2 metals (M1 & M2) Cdg	pGpG
monomeric ***FimX-EAL (3HV8)*** ***FimX-EAL (4FOJ)*** ***FimX-EAL (4FOU)*** ***FimX-EAL (4FOK)*** ***FimX-EAL (4F3H)*** ***FimX-EAL (4F48)*** ***LapD-EAL (3PJW)*** ***LapD-EAL (3PJX)*** ***LapD-EAL (3PJU)*** dimeric ***LapD-EAL (3PJT)***	**monomeric** PA3825-EAL (4Y9M) DosP-EAL (4HU3) *LapD-EAL (3PFM)* *FimX-EAL (3HV9)* **dimeric** YkuI (2BAS) *YdiV-EAL (3TLQ)* Imo0131-EAL (4Q6J) MorA-EAL (4RNJ) MorA-EAL (4RNI) *FimX-EAL (4AG0)*	**monomeric** **Mg^{2+}** CC3396-EAL (3S83) **Ca^{2+}** PA3825-EAL (4Y8E) **dimeric** **Mg^{2+}** TBD1265-EAL (2R6O) **tetrameric** **Mg^{2+}** RocR (3SY8)	**monomeric** **Mg^{2+}** MorA-EAL (4RNH) **dimeric** **Ca^{2+}** Blrp1_B (3GFX) YkuI_AB (2W27)	**monomeric** **Mg^{2+}** PA3825-EAL (4Y9Q) MucR-EAL (xxxx) **Ca^{2+}** CC3396-EAL (4HJF) **dimeric** **Mg^{2+}** TBD1265_EAL (3N3T) **Ca^{2+}** PA3825-EAL (4Y9P) Blrp1_A (3GFX) Blrp1 (3GG1) YahA (4LJ3) **Mn^{2+}** Blrp1 (3GFZ) Blrp1 (3GG0)	**monomeric** **Mn^{2+}** PA3825-EAL (xxxx) **dimeric** **Mg^{2+}** CC3396-EAL (3U2E) **no metal** *FimX-EAL (4AFY)*

Superposition of the ternary complexes of co-crystallised *PA*3825-EAL with CdG and pGpG in the presence of Mg^{2+} and Mn^{2+}, respectively, present with a near identical conformation in the catalytic pocket, with the exception of M3 in the pGpG bound structure (Fig. 6.29a). The only other structure to present with M3 is the unpublished pGpG-bound *CC*3396-EAL (PDB code: 3U2E) (Fig. 6.29a). *CC*3396-EAL was crystallised at pH 8 in the presence of magnesium so metals have been modelled as Mg^{2+} ions by the authors. Each Mg^{2+} ion coordinates to one of the available oxygen atoms from the hydrolysed phosphate group of pGpG (Fig. 6.26b and 6.26d). M1 and M2 sites correspond to previously observed M1 and M2 sites in other bimetallic EAL domain structures (Fig. 6.29a), with aninter-metal distance of 4.4 Å. The position of *PA*3825-EAL's M3 site corresponds with *CC*3396-EAL, as shown in figure 6.29b. However, the coordination of M2 and M3 varies between structures with M2 completely absent in *PA*3825-EAL. Instead, Glu217, a liganding residue of M2 when present, coordinates to the free oxygen from pGpG's phosphate (Fig. 6.26a and 6.26c). In the case of M3, no residues are involved in the octahedral coordination for *CC*3396-EAL. The corresponding Asp160 and Asp161 residues of *PA*3825-EAL are replaced with the O3' oxygen of the

sugar ring and an additional water, WC6 (Fig. 6.26b and 6.26d). Despite these differences in M3 coordination between structures both are coordinated to the two free oxygen atoms involved in forming the phosphoester bond of CdG prior to hydrolysis. The availability of this additional oxygen donor from pGpG may explain why the M3 site has only been captured in product bound structures. This novel M3 site is ideally placed to allow formation of a substrate transition state, suggesting EAL domains may possess a functional tri-metallic catalytic centre as opposed to previously suggested bimetallic centres.

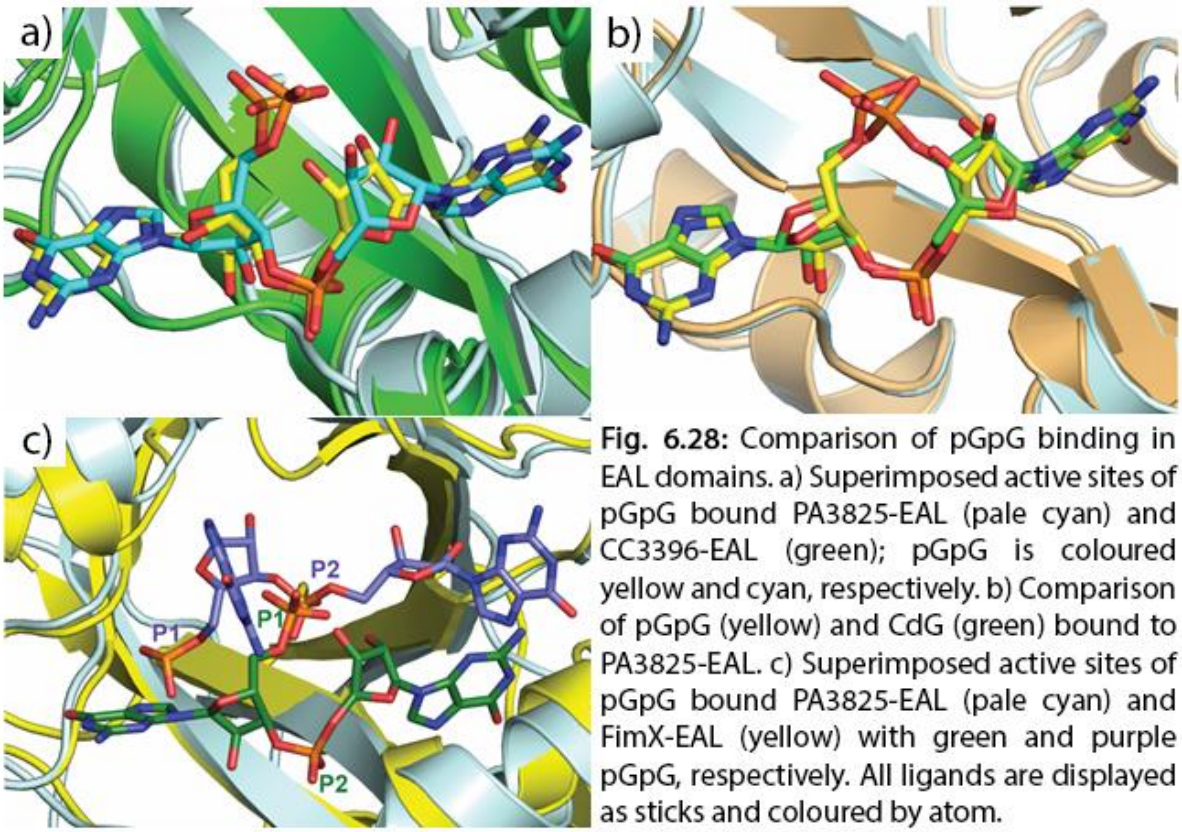


Fig. 6.28: Comparison of pGpG binding in EAL domains. a) Superimposed active sites of pGpG bound PA3825-EAL (pale cyan) and CC3396-EAL (green); pGpG is coloured yellow and cyan, respectively. b) Comparison of pGpG (yellow) and CdG (green) bound to PA3825-EAL. c) Superimposed active sites of pGpG bound PA3825-EAL (pale cyan) and FimX-EAL (yellow) with green and purple pGpG, respectively. All ligands are displayed as sticks and coloured by atom.

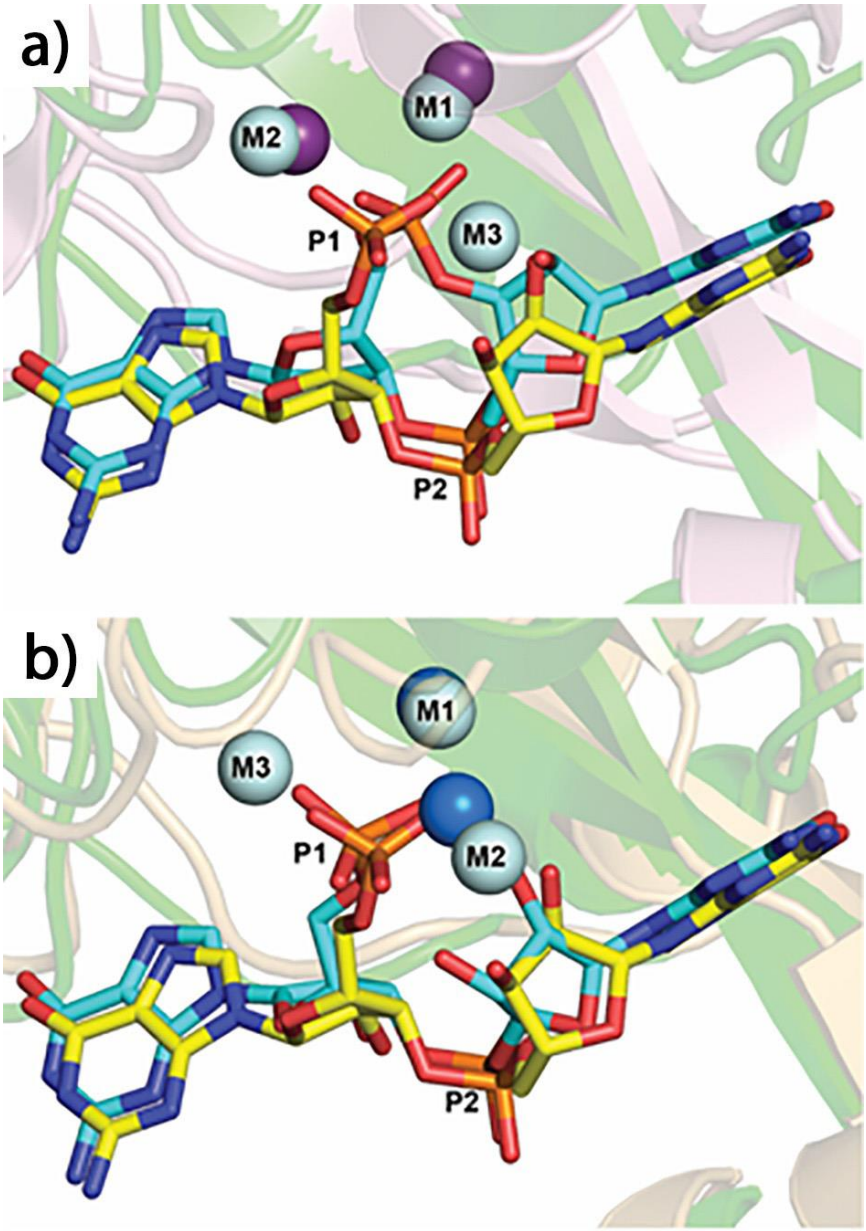


Fig. 6.29: Comparison of metal binding sites in pGpG and CdG bound EAL domains. a) Superposition of pGpG (yellow) bound to trinuclear magnesium-loaded *CC*3396-EAL (pale cyan and green, respectively) and CdG (cyan) bound to trinuclear magnesium-loaded *CC*3396-EAL (purple and pink respectively). b) Superposition of pGpG (yellow) bound to trinuclear magnesium-loaded *CC*3396-EAL (pale cyan and green, respectively) and pGpG (cyan) bound to manganese-loaded *PA*3825-EAL (blue and light orange, respectively).

6.3.5. MucR-EAL CdG-bound Structure and the Putative Inhibitory Dimerisation Interface

In addition to *PA*3825-EAL in various states a high resolution structure of Mg^{2+}-loaded CdG-bound MucR-EAL was obtained. Data were collected and phased by the Tews group at the University of Southampton with refinement and validation of the structure performed for comparison with *PA*3825-EAL at the University of Liverpool/ Diamond light Source. Data processing, refinement and validation statistics are reported in Table 6.A1. The αβ(βα)7 TIM-barrel fold of MucR-EAL (Fig. 6.30) is consistent with that of *PA*3825-EAL and other published structures including BlrP1 (Barends et al., 2009), YahA (Sundriyal et al., 2014), FimX (Robert-Paganin et al., 2012) and RocR (Chen et al., 2012). Both metal sites are fully occupied and coordination is consistent with *PA*3825-EAL and other previously published bimetallic EAL domain structures. The M1 site is octahedrally coordinated by Glu65, Asn124, Asp156, Asp186, free oxygen from the scissile phosphate (P1) of Cdg and the in-line attacking water molecule, WA1. M2 is also octahedrally coordinated by Asp186, Asp187, Glu243, a P1 free oxygen and three waters, WA1, W2 and W3. As seen in other bimetallic EAL domains Asp186 bridges the two metal ions as a carboxylate fork (Fig. 6.31).

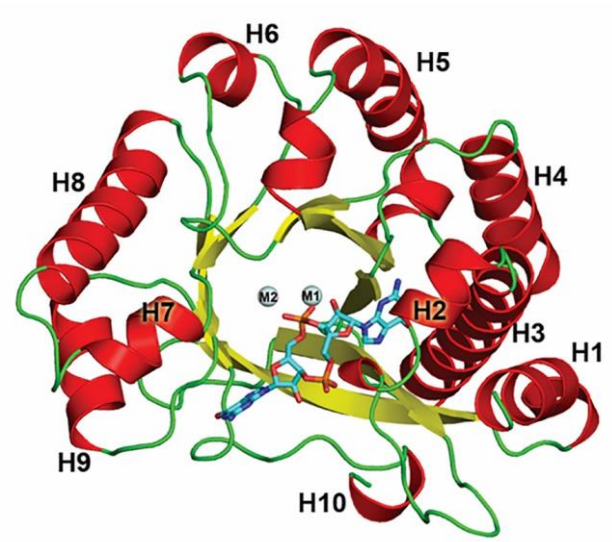


Fig. 6.30: Crystal structure of the Mg-loaded MucR-EAL monomer shown in cartoon form with α-helices in red, β-strands in yellow and loops in green. CdG is shown as sticks, coloured in cyan and by element. Catalytic Mg^{2+} ions, labelled M1 and M2, are displayed as spheres and coloured pale cyan. Helices have been numbered from 1-10.

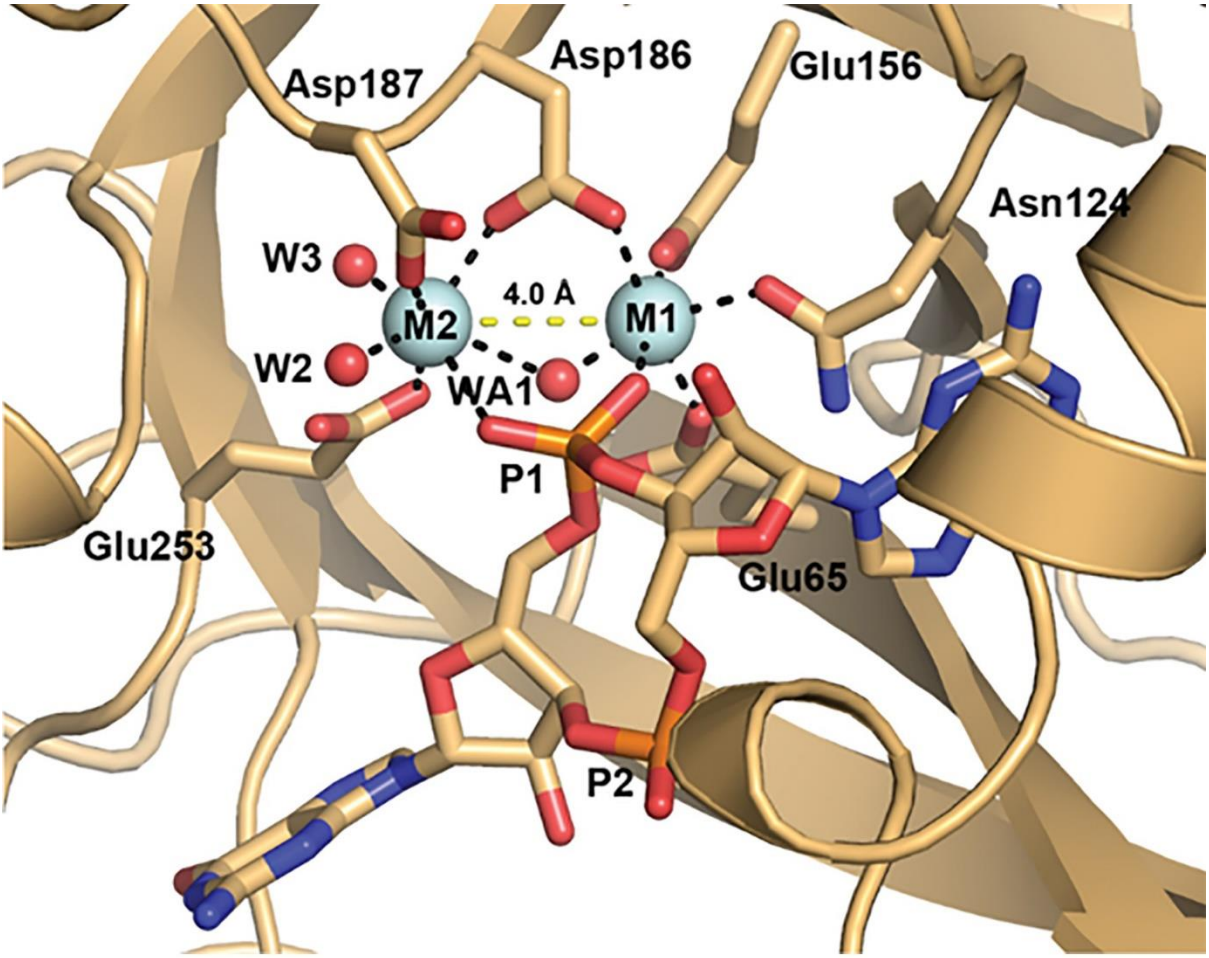


Fig. 6.31: A detailed view of the bimetallic centre of magnesium-loaded CdG-bound MucR with protein/ligand-metal and inter-metal distances shown as black and yellow dashed lines, respectively. Protein-metal liganding residues and CdG are shown as sticks and coloured by atom type. Mg^{2+} ions and water molecules are shown as cyan and red spheres, respectively.

Crystallisation of MucR was performed using purified protein with no co-crystallisation or soaking with CdG, however, the resulting MucR-EAL structure has a fully occupied catalytic pocket with two magnesium ions coordinating a CdG molecule; in agreement with the observed absorption at 270 nm suggesting bound nucleotide. Although earlier enzymatic assays confirmed MucR's inactivity in solution, MucR is theoretically an active PDE EAL at the purification pH based on the presence of conserved catalytic residues in its sequence (Fig. 6.17). Structural comparisons of CdG-bound MucR-EAL and CdG-/pGpG-bound *PA*3825-EAL show no significant differences that could explain this discrepancy in MucR's catalytic activity. As eluded to earlier, the only difference between the structure of MucR-EAL and *PA*3825-EAL is that MucR-EAL is completely dimeric while

*PA*3825-EAL is predominantly monomeric based on SEC-MALS data. However, the mechanism by which dimerisation of MucR monomers would cause inhibition of PDE activity is unclear from structural data. Recently published work on the influence of allosteric changes on Blrp1 activity through EAL domain dimerisation (Winkler et al., 2014) may shed slight on the inhibition of MucR. Winkler *et al* 2014 demonstrated that Blrp1 EAL activity is coordinated through the EAL dimerisation interface; particularly the role of the "compound helix", two short helices from each monomer, in metal coordination in the active site. Upon induction with light, Blrp1 undergoes a signalling cascade originating in the $\alpha1$-$\beta2_B$ and region $\beta5_B$ which, depending on the metal coordination, can affect $\beta4$-$\beta5_B$, $\alpha5$- $\alpha4_B$ loops, $\beta4_B$ and EAL dimerisation (Winkler et al., 2014). These light-induced structural changes cause a clam-shell like opening and closing action between EAL dimers governed by the "compound helix"; an evolutionarily conserved dimerisation interface in catalytically active EAL domains, and some inactive domains (Barends et al., 2009, Tchigvintsev et al., 2010, Minasov et al., 2009, Tarnawski et al., 2013, Chen et al., 2012). The dimerisation interface of MucR-EAL is highly conserved with Blrp1 (PDB code 3GG0) (Fig. 6.32), with the long helix 8 accounting for the main interactions. However, differences arise in the position of the short alpha helices that form the compound helix, circled in red in figure 6.33. In Blrp1 both helices run parallel to the long helix 8 (Fig. 6.32b, pink) while MucR shows both helices pointing away from helix 8 at approximately 45° (Fig. 6.32a).

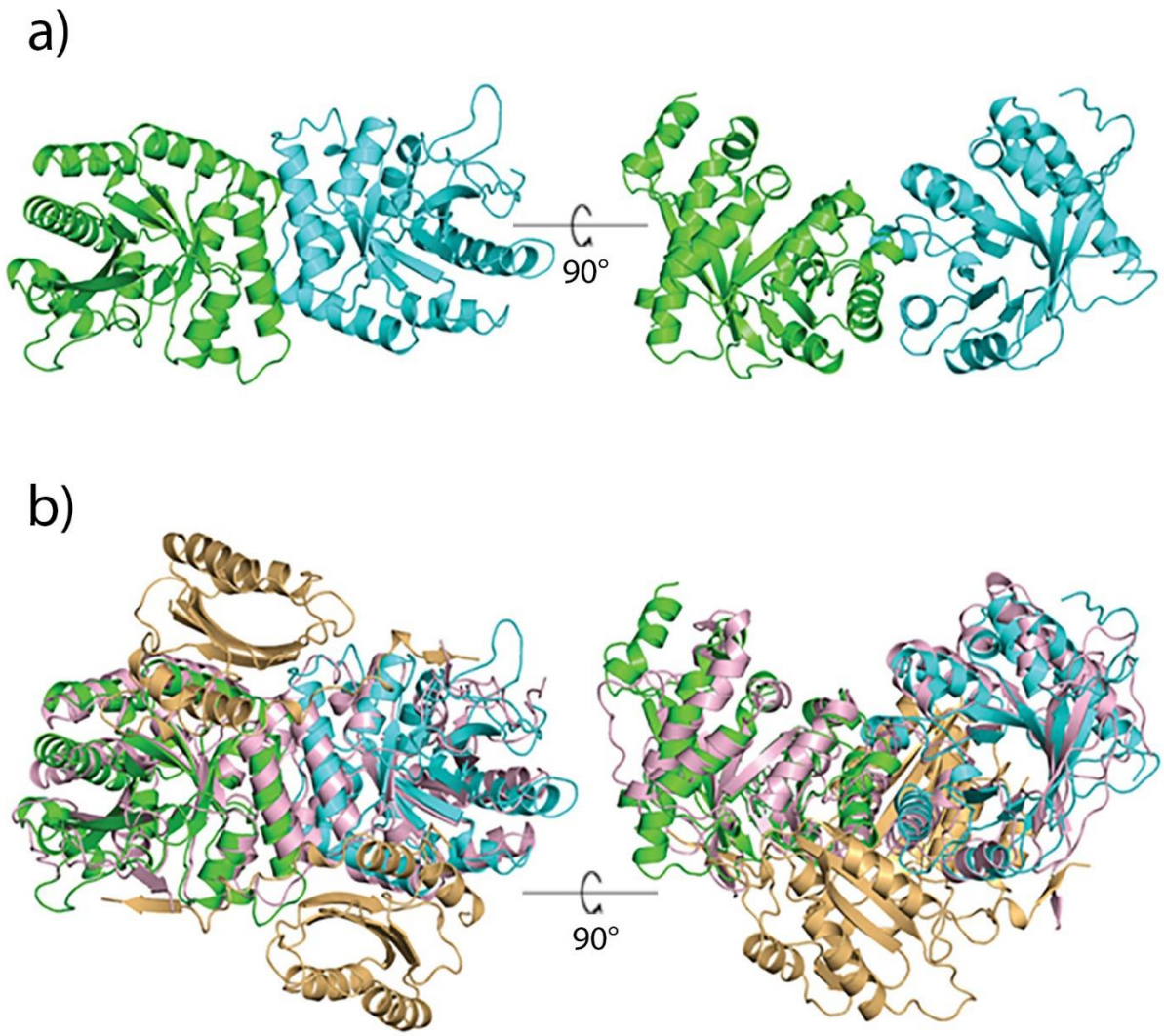


Fig. 6.32: Comparison of MucR and Blrp1 cyrstal structure dimerisation interfaces. a) Cartoon representation of the MucR dimer with monomers coloured green and cyan showing the helix 8 dimerisation interface. b) Superposition of MucR-EAL and Blrp1 demonstrating the same helix 8 mediated dimerisation interface. MucR is coloured as in (a) and Blrp1-EAL domains are coloured light pink with BLUF domains in light orange.

Despite the conserved dimeric conformation, sequence identity between MucR and Blrp1 is low at only 24%; while MucR and *PA*3825-EAL have a much higher sequence identity at 38% (Fig. 6.33a). Superposition of a *PA*3825-EAL monomer with MucR-EAL chain A show significant differences in helix 8 including His198Ala and Lys209Thr in *PA*3825-EAL which may suffice in preventing dimerisation in solution as observed in SEC-MALS experiments despite high sequence identity (Fig. 6.33b). If this dimeric conformation that is conserved throughout various EAL domains (Barends et al., 2009, Tchigvintsev et al., 2010, Minasov et al., 2009, Tarnawski et al., 2013, Chen et al., 2012)

does indeed convey an inhibitory effect, as seen in MucR-EAL, it may explain why Blrp1 crystal soaking experiments at pH 9 in the presence of Mn^{2+} coordinate CdG as opposed to the expected pGpG based on *in vitro* experiments (Barends et al., 2009).

6.4. Conclusions

EAL domains are crucial in the organisation of bacterial life cycles by regulating intracellular levels of the pivotal secondary messenger CdG. EAL domains are capable of coordinating pairs of calcium, magnesium or manganese ions through two highly conserved Asp residues on loop 6 (Barends et al., 2009, Sundriyal et al., 2014, Schmidt et al., 2005). Of these metal complexes; Mn^{2+}-loaded are significantly faster enzymes than the Mg^{2+}-loaded equivalents while calcium ions completely inhibit PDE activity (Sundriyal et al., 2014, Schmidt et al., 2005). Inhibition by calcium is arguably the result of Ca^{2+} acting as a weaker Lewis acid than Mg^{2+} and Mn^{2+} (Irving and Williams, 1953), preventing activation of the in-line attacking water to hydrolyse the phosphoester bind in CdG. It is not uncommon for metal ions to act as Lewis acids in catalysis for various reactions including activation of electrophiles (carbonyl, nitrile and phosphoryl groups), activation of leaving groups (alkoxide ions, amide ions, sulphur derivatives and halides), activation of ambient acids (water, alcohols and amines) and nucleophilic attack via metal bound hydroxide ions at carbonyl, nitrile and most importantly in the case of PA3825-EAL, phosphoryl groups (Suh, 1992). Here I have demonstrated *in vitro PA*3825-EAL converts CdG to pGpG in the presence of magnesium and/or manganese, with reaction rates considerably faster in the presence of manganese while calcium completely inhibits PDE activity. Reaction rates decrease with more acidic pH, in agreement with other EAL domain work (Barends et al., 2009).

Mg^{2+}-CdG-*PA*3825-EAL was co-crystallised at the inactivating pH of 7.1. This structure shows considerable similarity to the ternary complex of MucR-EAL and TBD1265-EAL (Tchigvintsev et al., 2010), the only other Mg^{2+}-loaded substrate bound EAL domain structures in the PDB (Table 6.3). Both structures were also obtained at an inactivating pH of 6.5. Although the catalytic pocket and

substrate orientation show significant similarities between all three proteins, the inter-metal distances between M1 and M2 vary considerably; with values of 4.0 Å, 4.4 Å and 4.2 Å for MucR-EAL, TBD1265-EAL and *PA*3825-EAL, respectively. This supports the idea proposed by Barends et al. (2009), that inter-metal distances may play a role in regulation of PDE activity, as all inhibited Mg^{2+} structures have a larger inter-metal distance than that observed in their highly active Mn^{2+}-loaded CdG bound Blrp1 structure.

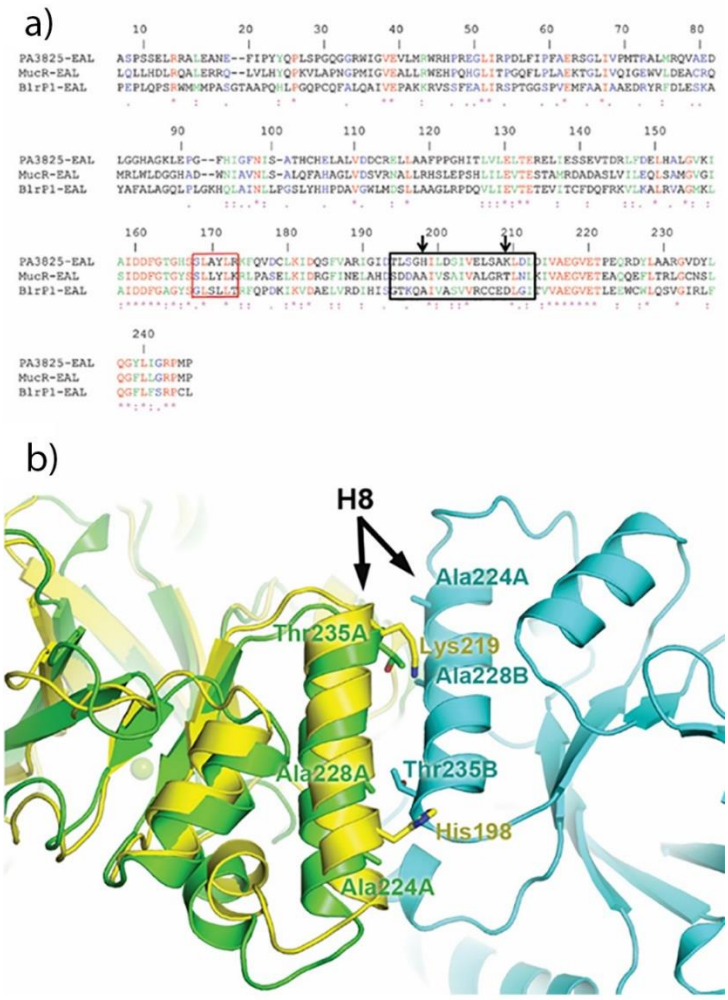


Fig. 6.33: Naturally occurring mutations between PA3825-EAL and MucR-EAL on helix 8 and their effect on dimerisation. a) Multiple sequence alignment of PA3825-EAL, MucR and Blrp1 EAL domains with helix 8 highlighted by the black box. The "compound helix" formed in Blrp1 dimers is highlighted by the red box. b) A detailed view of the MucR dimerisation interface through helix 8, with monomeric PA3825-EAL superimposed to chain A of the MucR dimer. His198 and Lys209 of PA3825-EAL may disrupt this domain-domain interaction compared to un-obtrusive Ala224 and Thr235 in MucR-EAL. PA3825-EAL is coloured in yellow and the MucR-EAL dimer chains A and B are coloured green and cyan, respectively.

Early studies into the reaction mechanism of EAL domains were performed through mutation studies on RocR paired with structural studies of CdG bound YkuL. Enzyme activity was tested against various mutations involved in metal binding, although many mutations saw a reduction in binding affinity for magnesium and activity compared to the wild type, only E352 proved absolutely essential for activity; remaining inactive even in the presence of 0.5 M Mg^{2+} (Rao et al., 2008). This led to a single metal general base catalyst mechanism with the essential glutamate residue acting as the general base to activate the in-line water for nucleophilic attack (Rao et al., 2008, Minasov et al., 2009). This model has since been expanded to a two metal catalytic mechanism in the light of CdG bound structures in the presence of pairs of Mg^{2+}, Mn^{2+} or Ca^{2+}; this suggests the metals act as a Lewis acid to activate the in-line attacking water (Tchigvintsev et al., 2010, Schmidt et al., 2005, Tamayo et al., 2005, Barends et al., 2009). The pGpG-bound structures presented in this work have revealed a new level of complexity in EAL domains with the observation of a third metal site in the EAL domain which requires further investigation.

A recent study has expanded further on the regulation of EAL domains through the effects of allosteric regulation of domain-domain dimerisation on PDE activity (Winkler et al., 2014). This may explain the presence of CdG in the crystal structure of Mn^{2+}-loaded Blrp1 at pH 9 (Barends et al., 2009), with conversion of CdG being prevented by allosteric effects of the dimer *in crystallo* restricting the protein or substrate to prevent formation of the M3 site and transition state activation by the third metal to allow PDE activity. Building on this, two EAL domains known to possess PDE activity as part of their physiological roles have failed to be crystallised with pGpG under conditions favourable to catalysis, Blrp1 (Barends et al., 2009) and MucR-EAL (this work). Although Blrp1 and MucR-EAL show a low sequence homology both share a conserved dimeric interface in crystal structures (this work). On the other hand, pGpG is clearly seen in the close homologue, *PA*3825-EAL, which is unable to form a homodimer due to mutations in the dimerising helix. The evidence for a catalytic inhibitory mechanism in EAL domains through dimerisation requires further study and mutation studies on *PA*3825-EAL's dimerisation helix may shed light on this observation.

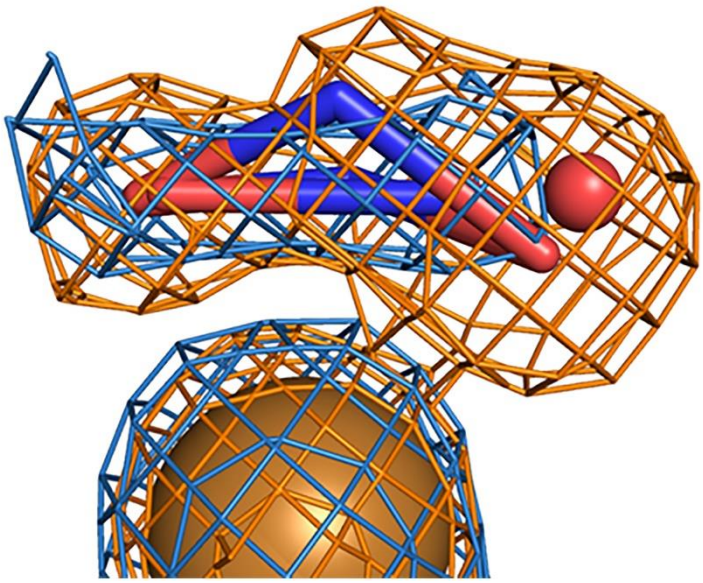
However, PDE regulation by such a means would fit well with the physiological distribution of EAL domains, which are regularly found adjacent to a diguanylate cyclase GGDEF domain; known to require dimerisation for conversion of two GTP molecules to CdG (Paul et al., 2007). Therefore a DGC-PDE coupled protein able to switch on activity of the GGDEF domain, while at the same time inhibiting the PDE activity of an EAL domain, through dimerisation and *vice versa*, presents an efficient means to control synthesis and degradation of CdG and regulate biofilm formation.

6.4. Appendix

Table 6.1A: Crystallization, crystallographic data collection and structure refinement statistics. Values in parentheses refer to the highest resolution bin.

Structure	**MucR-EAL**
Data Collection	
Space Group	$P12_11$
Cell Axes (Å)	a = 46.4, b = 116.1, c = 52.1
Angles (°)	β = 102.5
Beamline	I04
Wavelength (Å)	0.9795
Resolution (Å)	45.27-2.27 (2.33-2.27)
Unique Reflections (#)	24883 (1831)
Redundancy	6.9 (6.9)
R_{pim} (%)	6.9 (27.6)
I/σ(I)	11.3 (2.9)
Completeness (%)	99.9 (99.8)
Refinement	
Molecules/AU	2
R_{work}/R_{free} (%)	20.1/25.4
Rmsd	
Bond Length (Å)	0.006
Bond Angles (°)	1.853
B Factors (Å^2)	
Average B factor	33
Ramachandran Statistics (%)	
Favoured Regions	95.2
Allowed Regions	4.6

Chapter 7 – Serial Data collection of a Nitrite Reductase Crystal: A Structural Movie of Nitrite Reduction



7.1. Introduction

Achromobacter cycloclastes is an aerobic Gram-negative straight rod shaped bacterium with several flagella. It belongs to the genus *Achromobacter* in the family *Alcaligenaceae* of the order *Burkholderiales* and is commonly found in fresh and marine water, soils and as a cell culture contaminant in laboratories (Gray et al., 2010). *A. cycloclastes* plays a key role in the nitrogen cycle, along with many other bacteria, which exclusively fix atmospheric dinitrogen to metabolically active forms to allow production of the amino acids, proteins and nucleic acids essential for life (Suzuki et al., 2000). The Nitrogen Cycle consists of six reactions, (1) the fixation of dinitrogen gas (N_2) to ammonium (NH_4^+); (2) aerobic ammonium oxidation to hydroxylamine (NH_2OH) by bacteria and archaea; (3) aerobic oxidation of nitrite (NO_2^-) to nitrate (NO_3^-); (4) denitrification of nitrite to nitric oxide (NO); (5) anaerobic ammonium oxidation of ammonium and NO to produce hydrazine (N_2H_4) and (6) dissimilatory nitrate and nitrite reduction to ammonium (Fig. 7.1) (Jetten, 2008). This work focuses on the mechanism by which nitrite reductases cleave nitrite to NO as part of denitrification.

Denitrification is an important step in the nitrogen cycle whereby dissimilatory reduction of nitrate or nitrite via gaseous N-oxides produces dinitrogen. This process has widespread effects on agriculture through loss of essential nitrate and nitrite and on the environment; principally increasing the effects of global warming by transferring excess nitrogen from hydrological systems to the atmosphere as N_2O and NO_x (Zumft, 1997, Suzuki et al., 2000, Bouwman et al., 2013). Production of NO by bacteria (coupled with proton translocation and ATP synthesis (Zumft, 1997)) occurs in low oxygen environments and in the presence of nitrite, increasing the expression levels of the nitrite reductase enzymes that perform the first committed step in the dissimilarity denitrification pathway (Boulanger et al., 2000, Starkenburg et al., 2008). Nitrite reductases are found in two structurally distinct forms defined by their catalytic centres, which contain either heme cd_1 prosthetic groups or Cu centres. Three quarters of tested bacterial strains have been found to possess the tetraheme

protein cytochrome *cd₁* (*cd₁*-NiR) with the remaining quarter expressing copper nitrite reductases (CuNiRs). Both NiRs are never found within the same cell (Zumft, 1997).

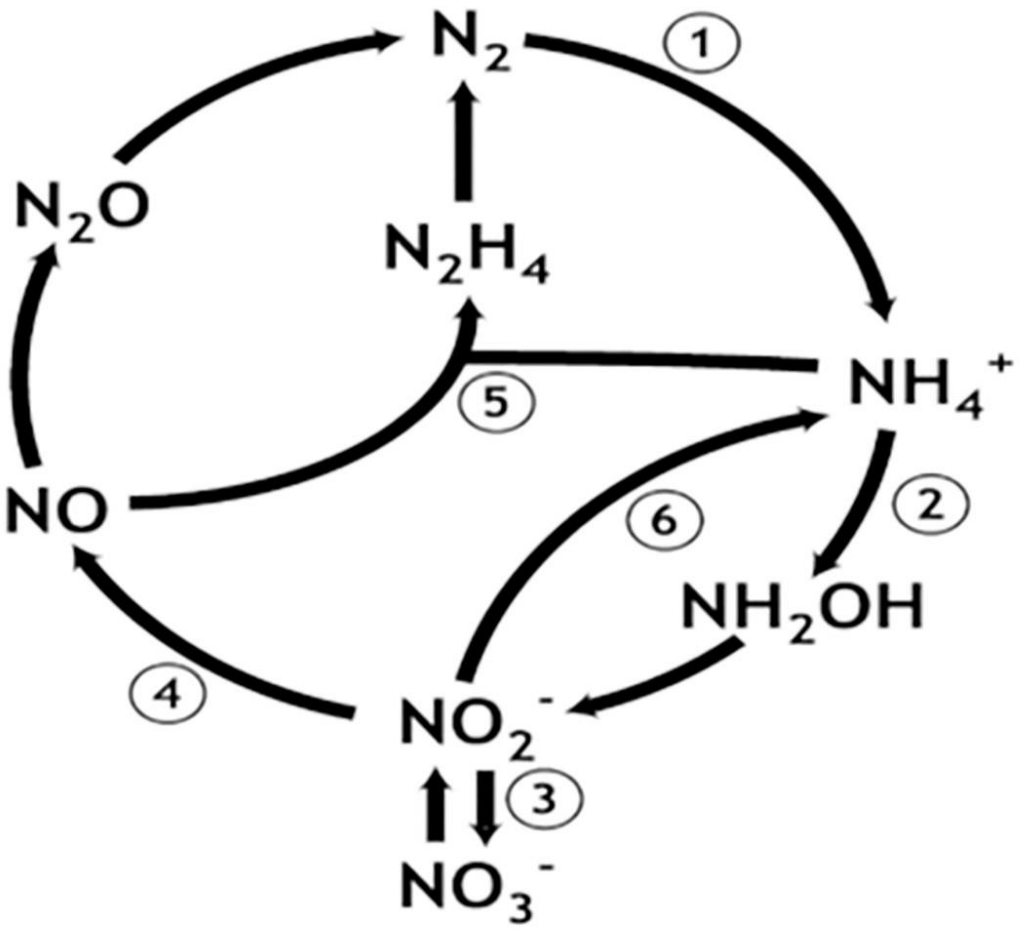


Fig. 7.1: A schematic of the reactions of nitrogen cycle, (1) fixation of dinitrogen gas; (2) aerobic ammonium oxidation to hydroxylamine; (3) aerobic oxidation of nitrite to nitrate; (4) denitrification of nitrite to nitric oxide; (5) anaerobic ammonium oxidation of ammonium and (6) nitric oxide to produce hydrazine and dissimilatory reduction to ammonium

cd₁-NiRs are respiratory nitrite reductases made up of homodimers of 60 KDa subunits with a heme C and heme D_1 domain. The heme D_1 domain forms a β-propeller structure consisting of 8 blades with heme D_1 positioned close to the subunit interface. The heme C domain on the other hand is predominantly helical with heme C positioned approximately 20 Å away from heme D_1. However, this study focuses on the structurally distinct CuNiRs so *cd₁*-NiRs will not be discussed in further detail.

7.1.1. Copper Nitrite Reductases

CuNiRs can be separated into three classes; (1) the type one blue copper proteins; (2) type two non-blue (green) copper proteins and (3) type three binuclear and EPR inactive copper proteins (Malkin and Malmström, 2006). The difference in colour stems from the T1Cu with blue enzymes showing absorbance in the 460 nm range and green enzymes with absorbance maxima around 460, 595 and 700-750 nm. T1Cu liganding residues between blue and green enzyme are the same (Adman et al., 1995), the only major difference being a shift in the His-Cu-Met angle by 17° which may affect the electronic structure (Dodd et al., 1998, LaCroix et al., 1996). An alternate theory suggests Tyr in green enzymes being replaced with Thr in blue enzymes causes a shift in the polypeptide backbone around the Met T1Cu liganding residue causing the colour change (Inoue et al., 1998). The structure of CuNiRs differs significantly from cd_1-NiRs with all CuNiRs showing a highly conserved trimeric structural arrangement (Fig. 7.2a) with each two domain subunit forming two "Greek key" β-barrel folds with each barrel stacked onto each other (Fig. 7.2b) (Godden et al., 1991, Zumft, 1997). Each subunit contains a T1Cu and T2Cu site. The tetrahedral type one copper (T1Cu) site buried in the N-terminal domain of the protein is formed from four adjacent residues in domain one ($His95_{Ac}$, $Cys136_{Ac}$, $His145_{Ac}$ and $Met150_{Ac}$) (Fig. 7.2c). The pseudohedral type two copper (T2Cu) ligand-binding active site is formed from three histidine residues ($His100_{Ac}$, $His135_{Ac}$ and $His306B_{Ac}$), two from domain one and one from the C-terminal end of domain two from an adjacent monomer in the trimer (Fig. 7.2c), and a forth distal water ligand (Godden et al., 1991, Kukimoto et al., 1994, Zumft, 1997). The overall mechanism by which CuNiRs reduce nitrite involves the transfer of electrons from physiological electron donors, azurin for blue CuNiRs and pseudo-azurin for non-blue CuNiRs, to the T1Cu which is then transferred to the T2Cu approximately 12 Å away at the subunit interface (Hough et al., 2005, Kukimoto et al., 1994). This mechanism has been investigated for a number of bacterial CuNiRs through a variety of site directed mutagenesis studies paired with, crystallography, EXAFS and spectroscopic and other enzymatic methods.

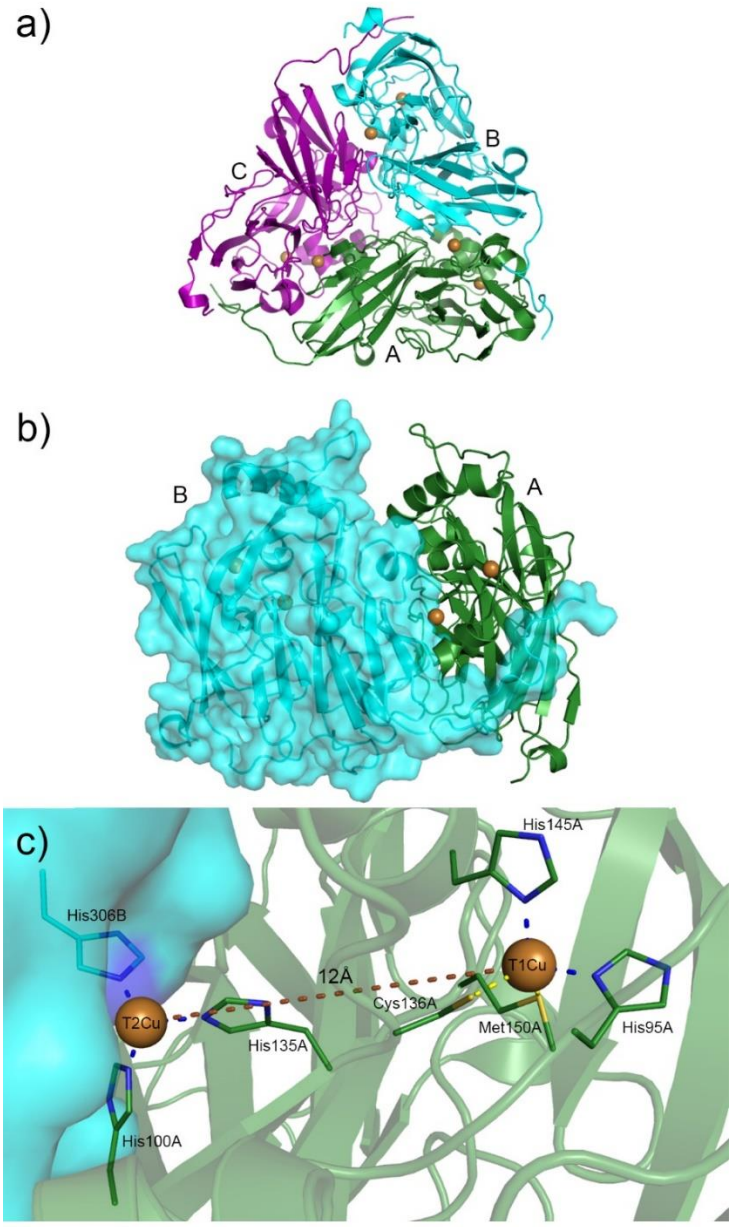


Fig. 7.2: The conserved structural arrangement of copper nitrite reductases. a) The functional trimer represented as a cartoon with copper as spheres and coloured green, cyan and purple for monomer A, B and C respectively. b) The dimer interface with monomer A as a cartoon and monomer B as a surface representation with monomers coloured as in (a). The T1Cu electron donor and T2Cu active site with lignding residues as sticks coloured by monomer as in (a). Figure made with PDB model 2WBI.

Nishiyama et al. (1992) performed site directed mutagenesis on pseudo-azurin from *Alcalegenes faecalis* replacing Pro80 with Ala to alter binding of one of the four T1Cu liganding residues. Pro80Ala mutation caused a marked increase in pseudo-azurin's reduction potential, 139 mV higher than the native, and a considerable decrease in electron transfer to NiR. The crystal

structure of the Pro80Ala mutant revealed a previously unseen water molecule bound to the Cu atom, suggesting increased reduction potential through greater solvent accessibility to the Cu atom (Nishiyama et al., 1992). This study gave the first insight into the structural elements involved in electron transfer in NiRs. This work was expanded by crystal structures of *Achromobacter cycloclastes* NiR (*AcNiR*) (Godden et al., 1991) and the work of Libby and Averill (1992) and Abraham et al. (1993) which proved the presence of both the T1Cu and the active site T2Cu in nitrite reductases. Following this, Kukimoto et al. (1994) introduced Met150Glu and His135Lys mutations to *Alacaligenes faecalis* nitrite reductase (*Af*NiR) altering T1Cu and T2Cu binding, respectively, to test copper binding and nitrite-reducing activity. Met150Glu mutations resulted in the loss of the T1Cu, as determined by optical absorption spectroscopy and atomic absorption spectroscopy, low activity using methyl viologen as an electron donor and no activity with pseudo-azurin as the electron donor. This strongly suggests the T1Cu is not directly involved in nitrite reduction but serves as the electron acceptor from pseudo-azurin which is then transferred to the T2Cu (Kukimoto et al., 1994). His135Lys mutations at the T2Cu active site showed weaker copper binding but both T1Cu and T2Cu were present. However, nitrite reducing activity was completely abolished as a result of the altered T2Cu environment (Kukimoto et al., 1994).

7.1.2. The Catalytic Mechanism

With an increasing number of crystal structures of both green and blue CuNiRs (Adman et al., 1995, Dodd et al., 1997, Dodd et al., 1998, Godden et al., 1991, Kukimoto et al., 1994, Murphy et al., 1997), as well as steady state kinetics (Abraham et al., 1997), EXAFS (Strange et al., 1995) and ENDOR (Olesen et al., 1998) data, an ordered mechanism by which nitrite binds the oxidised T2Cu site prior to the T2Cus reduction by the T1Cu was proposed (Strange et al., 1999). This study also identified a Cys-His link between T1Cu and T2Cu which acts as an internal electron transfer route and a His-Asp-His peptide stretch to signal the presence of nitrite at the T2Cu (Strange et al., 1999). Comparison of CuNiRs with similar zinc sites from superoxide dismutase, thermolysin, astacin and

carbonic anhydrase identified a proton abstracting group residue (Asp98$_{Ac}$) and a catalytic role for His255B$_{Ac}$ as a proton donor in catalysis (Fig. 7.3) (Murphy et al., 1995, Strange et al., 1995, Murphy et al., 1997, Strange et al., 1999). To further investigate the roles of these residues, D98N, H255D and H255N were produced in *Af*NiR for structural, biochemical and biophysical analysis (Boulanger et al., 2000).

All three mutants caused a significant reduction in the specific activity of *Af*NiR with a 94-, 99- and 112-fold decrease for D98N, H255D and H255N, respectively. The crystal structure of *Af*NiR-D98N shows a 1 Å displacement relative to the native Asp, preventing formation of a hydrogen bond with the water molecule at the T2Cu which has shifted by ~ 0.8 Å. This movement of the T2Cu-water in D98N from the native structure suggests a stabilising role for Asp98 in nitrite binding consistent with proton donation. The importance of Asp98 in substrate binding and maintaining the T2Cu geometry was investigated further through EPR spectroscopy and structural studies on D92N mutations in *Alcaligenes xylosoxidans* NiR (*Ax*NiR) (an Asp98 equivalent) which lead to T2Cu centres with higher redox potentials and a lack of hydrogen bonding to the T2Cu-water ligand (Prudêncio et al., 2001, Ellis et al., 2002). Both H255D and H255N lead to the appearance of a new water molecule between the mutant residue and Asp98 and H255N causes the T2Cu-water to be displaced by 1.4 Å from the native position (Boulanger et al., 2000). The consistent incorporation of an additional water to replace the lost proton donor for the Asp98-His255B bridging water suggests an important role for His255B as a proton donor in the water network. The importance of the hydrogen bonded water network was investigated further by introducing a H254F mutation in *Ax*NiR. Crystal structures revealed the loss of side chain nitrogen atoms from His254 caused a breakdown in the water network and the incorporation of zinc in place of the T2Cu, leading to a complete loss of activity (Ellis et al., 2002).

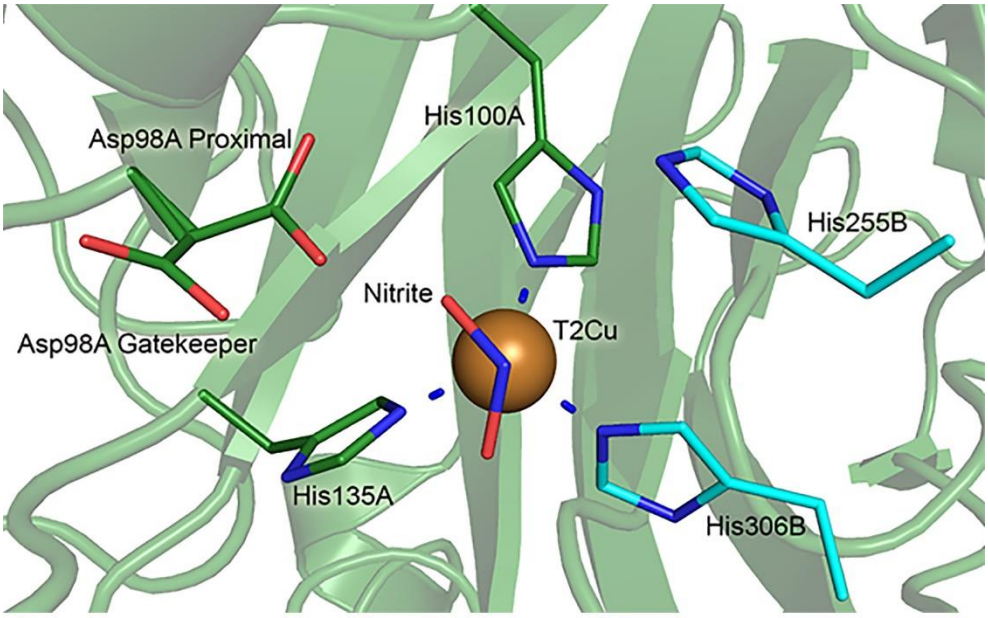
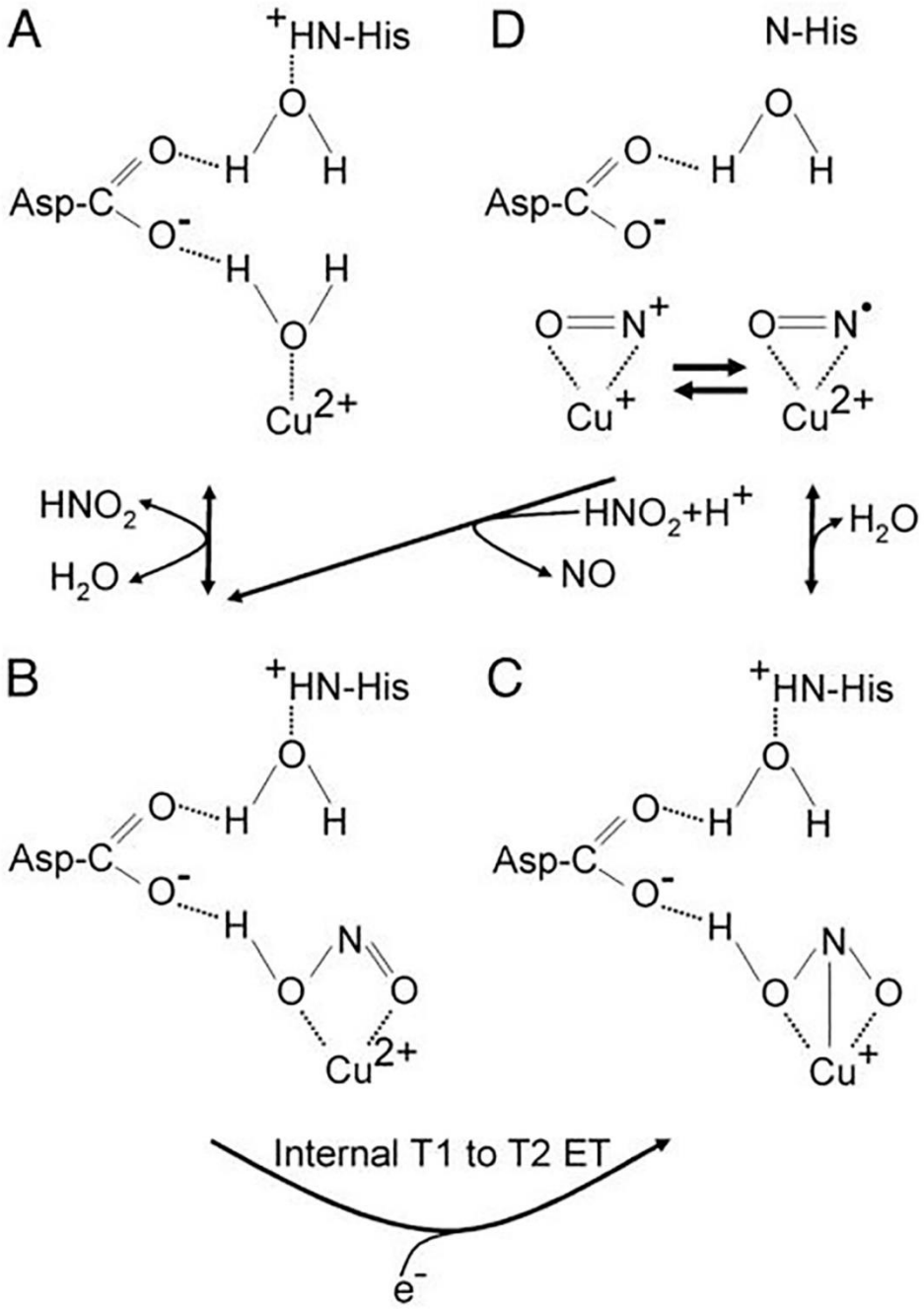


Fig. 7.3: The T2Cu active site of copper nitite reductases with copper liganding his residues (His100A, His135A and His306B), the dual conformation proton abstrating residue (Asp98A) in the proximal and gatekeeper positions and a second proton donor (His225B). All residues and ligands are represented as sticks with residues from the A monomer in green and the B monomer in cyan. Figure generated from *Ac*NiR model 2BWI.

Although the essential role of Asp98 and His255B (and their equivalents) has been demonstrated through mutagenesis studies discussed above, another residue, Ile257B, has been found to play a key role in substrate recognition and orientation of nitrite at the active site. Six Ile257 mutations were introduced into *Af*NiR (I257V, I257L, I257A, I257T, I257M and I257G) and nitrite soaked crystal structures solved. Native, I257V and I257L mutants all show bidentate binding of nitrite through the oxygen atoms of nitrite, with I257V actually producing a more active form of the enzyme by producing a shorter hydrogen bond between nitrite and Asp98 (Boulanger and Murphy, 2003). I257A, I257G, I257T and I257M mutations on the other hand lead to a radical reorientation of nitrite which bound in a monodentate orientation to the T2Cu, disrupting the essential Asp98-nitrite hydrogen bond (Boulanger and Murphy, 2003). Based on this it can be concluded that Asp98 and His255B alone are not sufficient for effective nitrite binding without Ile257B.

Following this work, atomic resolution structures of *Ac*NiR in the resting state, substrate bound and product bound states expanded on the role of Asp98 with the observation of alternate conformations suggesting a role in guidance of nitrite to the T2Cu as well as proton abstraction and formation/release of the product (Antonyuk et al., 2005, Tocheva et al., 2004). The appearance of multiple conformations pointing into the substrate entry pocket (gatekeeper conformation) and towards the substrate binding pocket (proximal conformation) (Fig. 7.3) with corresponding occupancies between the proximal conformation and nitrite at the T2Cu strongly suggests Asp98 plays a role in "hand-shaking" nitrite and guiding it to the T2Cu (Antonyuk et al., 2005). Based on these high resolution structures at each stage of the reaction an enzyme mechanism was proposed. Resting state NiR is observed with water bound to the T2Cu stabilised by a hydrogen bond from Asp98 in the proximal position which is hydrogen bonded to a protonated His255B from an adjacent monomer via the bridging water discussed earlier (Scheme 1a). Protonated nitrite then binds the T2Cu by displacing the water molecule and reforming the H-bond with Asp98 (Scheme 1b). Nitrite binding increases the mid-point potential of the T2Cu to facilitate electron transfer from the T2Cu (Hough et al., 2005) and proton transfer from Asp98 to nitrite (Scheme 1c) followed by donation of the second proton from the imidazolium group of His255 to cleave the O-NO bond to form NO and water (Scheme 1d) (Antonyuk et al., 2005).



Scheme 1 (Antonyuk et al., 2005)

Experiments on *Ax*NiR measuring changes in the phenol red spectrum (a pH indicator) during substrate turnover have shown approximately two protons are consumed over the course of the reaction (Brenner et al., 2009). Putative proton channels have been identified in *Ax*NiR and *Ac*NiR, one which connects the active site to the surface of the molecule via $Asn90_{Ax}$ and a second

which runs along the monomer-monomer surface via His254$_{Ax}$ (Dodd et al., 1998, Antonyuk et al., 2005). Disruption of the His254 pathways by H254F mutations caused zinc to bind in the active site but had no effect on catalytic activity when zinc is replaced with copper. On the other hand, mutation of Asn90 to serine caused a ~ 70% reduction in activity (Wijma et al., 2006, Leferink et al., 2011). Crystal structures of *Ax*NiR N90S, H254F and N90S-H254F mutants show disruptions in their respective proton channels. N90S and N90S-H254F mutants see a shift in Pro132 which moves away from the T1Cu coordinating His139 towards the mutated Ser90 residue, likely causing the 60 mV increase in the redox potential of the T1Cu (Leferink et al., 2011). Nitrite soaked structures of *Ax*NiR N90S show a more dramatic structural change from the native with Ala131 and Pro132 moving away from the T1Cu allowing Glu133 to move towards Ser90 and restore the proton delivery channel. Dramatic loss of activity and considerable structural changes upon nitrite binding suggest that the "Asn90$_{Ax}$" channel is the major provider of protons to the T2Cu active site but protons are still able to leak into the active site, as evident from the retention of some catalytic activity in N90S mutants (Leferink et al., 2011, Leferink et al., 2012b).

The electron transport between the T1Cu and T2Cu has been shown to be of paramount importance in the mechanism of CuNiR, whose metal site is otherwise ready to carry out the reaction (Moura et al., 2008). Of particular importance is nitrite binding prior to reduction of the T2Cu to increase the midpoint potential of the T2Cu to facilitate electron transfer from the T1Cu and drive catalysis (Hough et al., 2005). Control of this process has been attributed to the substrate sensor loop (His94$_{Ax}$-Asp92$_{Ax}$-His89$_{Ax}$) leading to electron transfer through a highly conserved Cys130$_{Ax}$-His129$_{Ax}$ bridge between the T1Cu and T2Cu (Strange et al., 1999). Introduction of mutants at the T1Cu (M144L$_{Ax}$, M144Q$_{Ax}$ and C130A $_{Ax}$) have been shown to disrupt the finely balanced set of redox potentials between azurin/pseudo-azurin, T1Cu and T2Cu. M144L mutations cause a 96 mV increase in redox potential, 31 mV higher than azurin and actually increasing the rate of catalysis through an increase in the positive ΔE value. M144Q on the other hand reduced the redox potential of *Ax*NiR by 132 mV rendering the protein completely inactive but still able to bind nitrite. C130A

mutation was much more disruptive, resulting in a completely empty metal site at the T1Cu centre and a mixture of nitrite bound Cu/Zn at the T2Cu centre. Together this study elegantly demonstrates loss of electron transfer from azurin to the T1Cu is the root cause of inactivity and azurin cannot transfer electrons directly to the T2Cu centre which could prematurely reduce the T2Cu before nitrite binding (Hough et al., 2005).

Despite this extensive structural knowledge of CuNiRs certain mechanistic aspects have been the subject of significant controversy, particularly whether nitrite binds the T2Cu prior to electron transfer (Adman et al., 1995, Hough et al., 2005, Murphy et al., 1997, Strange et al., 1999), binds after electron transfer (Abraham et al., 1997, Averill, 1996, Wasser et al., 2002) or if the two routes work in parallel (Averill, 1996, Wasser et al., 2002). Steady state kinetics data from CuNiRs at different pHs and in the presence of low and high nitrite fitted a random-sequential model where in low nitrite the T2Cu was reduced before nitrite binding. In high nitrite the T2Cu was reduced after binding, at pHs < 6 with 1 mM nitrite the T2Cu is reduced after nitrite binding and at pHs > 6 the T2Cu is reduced before nitrite binding (Wijma et al., 2006). The authors suggest this stems from the presence of different ligands bound at the T2Cu, in more alkaline conditions water is bound to the T2Cu which is easily displaced by nitrite while in more acidic conditions water is converted to OH- which is more difficult for nitrite to displace so the T2Cu is reduced first (Wijma et al., 2006). However, a study combining X-ray crystallography, online optical X-ray absorption spectroscopy (XAS) and optical spectroscopy on AxNiR crystals has since provided compelling evidence to the contrary in favour of an ordered mechanism in CuNiRs. Optical spectroscopy prior to exposure to X-rays shows T1Cu in the oxidised Cu(II) form. Following X-ray data collection with an estimated dose of 1.5×10^6 Gy the T1Cu had been reduced, as evident from the loss of the characteristic optical absorption peak at 595 nm (Hough et al., 2008b). Following a second data set on the same protein crystal an XAS spectrum was collected which showed no shoulder at 8985 eV characteristic of a reduced T2Cu centre (Strange et al., 1999, Wijma et al., 2006, Hough et al., 2008b). Analysis of various high resolution crystal structures (> 1.4 Å) also demonstrated an increase in the Cu-Cys bond

length of 0.09-0.13 compared to the distances determined for the oxidised T1Cu site as determined by XAS (Hough et al., 2008b). These data unambiguously demonstrate the mechanism of an ordered mechanism with gated electron transfer through binding of nitrite to the T2Cu centre.

Further study of the gated electron transfer mechanism using laser flash photolysis to rapidly reduce the T1Cu by reducing solvated electrons and other radicals. Initial reduction occurs with a κ_{obs} ~ 3000 s^{-1}, resulting in a colourless T1Cu, followed by a slower re-oxidation at ~ 370 s^{-1} representing redistribution of electrons to the T2Cu. For complete redistribution to occur, bound nitrite is required (Leferink et al., 2012b, Leferink et al., 2012a). In N90S mutated *Ax*NiR, rapid reduction of the T1Cu shows no subsequent recovery in the apo state and a much slower recovery compared to the native in the presence of excess nitrite. This corroborates the idea that nitrite binding increases the redox potential of the T2Cu to facilitate electron transfer (Strange et al., 1999, Olesen et al., 1998, Leferink et al., 2012a). Analysis of proton transfer with phenol red in native *Ax*NiR following laser flash photolysis gave transfer rates of 305 and 176 s^{-1} in the absence and presence of nitrite, respectively; while N90S mutants showed no transfer in the absence of nitrite and very slow transfer in the presence of nitrite (Leferink et al., 2012a). These results demonstrate tight coupling between electron and proton transfer in CuNiRs with a very similar dependence on nitrite binding. Leferink et al. (2012a) have also demonstrated that protonation occurs before electron transfer by chemically reducing *Ax*NiR with penanazine at pH 9, to give a colourless reduced T1Cu, which returns to its original blue colour and produces NO when the pH is reduced to 6.7. This shows that protonation, in the form of the pH drop, has triggered the electron transfer to the T2Cu to allow the production of NO.

Building on the evidence that the T1Cu is rapidly reduced by X-rays, this work seeks to introduce a sequential data collection technique whereby radiolysis by X-ray radiation initiates the electron transfer process to allow the catalytic mechanism of CuNiRs to be followed through high resolution crystallographic data to better understand the dynamic processes that the important

residues discussed above go through. This collection method produces a dose-dependent series of structures allowing a “structural movie” of catalysis to be made by X-ray induced radiolysis. Of paramount importance is that X-ray induced redox state changes occur at much lower X-ray doses than classical radiation damage. This potentially allows many (from tens to hundreds, depending on the system) of sequential structures to be determined while the crystal continues to diffract to high resolution.

X-ray radiation damage has been a long standing problem in structural biology (Garman, 2003a, Garman, 2010, Garman and Owen, 2006, Ravelli and Garman, 2006). However, the rapid reduction of redox centres has only become widely recognised recently, occurring before deteriorating diffraction resolution and damage to disulphide bridges, thiols and other side chains (Yano et al., 2005a, Beitlich et al., 2007, Antonyuk and Hough, 2011). Although these X-ray induced changes can complicate analysis of redox-active systems through mis-assignment of functional states or stages in catalytic mechanisms, limiting the X-ray dose can allow structures to be determined in specific redox states relevant to catalytic or other important protein mechanisms (Schlichting et al., 2000, Hough et al., 2008b). This work has been made feasible thanks to the recent introduction of fast shutterless PILATUS detectors (Rajendran et al., 2011a) at synchrotron sources which allow rapid measurement of large number of sequential datasets from the same region of a protein crystal at the low dose per dataset required. *Ac*NiR has been selected as a test sample as it has been shown to diffract to atomic resolution (Antonyuk et al., 2005) and is resistant to radiation damage. The development of streamlined automated data processing pipelines such as Xia2 (Winter et al., 2013) and easily scriptable software such as XDS (Kabsch, 2010a, Kabsch, 2010b) and AIMLESS (Evans and Murshudov, 2013) have made it possible to work with the terabytes worth of diffraction data that are generated by the sequential method of data collection used here. Such an approach is heavily reliant on using these automated software and computer clusters effectively.

7.2. Methods

7.2.1. Protein Production and Purification

An *Ac*NiR gene codon optimised for expression in *E. coli* cloned into a pet26b plasmid was purchased from Genscript and transformed into BL21 *E. coli* cells in the presence of 30 µg/ml kanamycin. LB cultures were grown with 2 mM $CuSO_4$ to an OD_{600} of 0.4-0.6 and induced with 2 mM IPTG at 18°C for 16 hours. Cultures were spun at 5000 rpm (Sorvall SLC6000 rotor) for 10 minutes at 4°C for 10 minutes and pellets resuspended in 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. Cells were lysed by sonication for 5 cycles of 30 seconds on and 30 seconds off and clarified by centrifugation at 16,000 rpm (Sorvall SS34 rotor) for 15 minutes at 4°C. Cleared lysate was dialysed for 3 hours against 2 mM $CuSO_4$ and 20 mM Tris-HCl pH 7.5 at 4°C. CuSO4 was removed by dialysis against 20 mM Tris-HCl pH 7.5 giving a distinct green colour.

Cleared dialysed lysate was loaded onto a homemade gravity flow hydroxyapatite column (Biorad, Hercules, California, USA) pre-equilibrated in 20 mM Tris-HCl pH 7.5 and washed with 20 ml of 20 mM Tris-HCl and 20 ml of 10 mM potassium phosphate. *Ac*NiR was eluted with increasing potassium phosphate from 10 mM to 150 mM by observing the green band of protein moving down the column. Eluted protein was concentrated and further purified by ammonium sulphate precipitation. 120 µL of protein was added to 240 µL of 20 mM Tris-HCl pH 7.5 and 40 µL of 100 mM sodium acetate pH 4.75. 25 µL of 4 M ammonium sulphate was added to the protein solution and left for 10 minutes at room temperature before pelleting precipitated protein by centrifugation at 14,000 rpm for one minute at room temperature. The supernatant was pipetted into a fresh Eppendorf tube and the process repeated until the pellet appeared green. Excess Ammonium sulphate was added to precipitate all *Ac*NiR protein which was pelleted by centrifugation and resuspended in 0.5 ml of 50 mM MES pH 6.5.

7.2.2. Crystallisation and NO_2/NO Soaking

Purified *Ac*NiR in 50 mM MES pH 6.5 was concentrated to 15-20 mg/ml as quantified by absorbance at 280 nm. Dark green regular trigonal pyramid shaped crystals in space group $P2_13$ were grown by hanging drop vapour diffusion at 19°C. 6 µL drops consisting of 3 µL protein solution and 3 µL 1.2-1.7 M ammonium sulphate and 100 µM sodium acetate pH 4.5 were equilibrated against 500 µL of 1.2-1.7 M ammonium sulphate and 100 µM sodium acetate pH 4.5 for several days. Small *Ac*NiR crystals were used as seeds to produce large single crystals by the same method described above.

Nitrite (NO_2) soaked crystals were produced by transferring large crystals into drops consisting of 2.4 M ammonium sulphate, 100 mM sodium acetate pH 5.5 and 100 mM sodium nitrite. Crystals were left for one hour at room temperature before flash cooling crystals in 100 mM nitrite and 3.5 M sodium malonate pH 5. Nitric oxide (NO) soaked crystals were produced by transferring crystals to drops containing 2.4 M ammonium sulphate and 100 mM sodium acetate pH 4.5 and equilibrating against 500 µL of 2.4 M ammonium sulphate, 100 mM sodium acetate pH 4.5 and 100mM sodium nitrite. Crystals were left at room temperature for several hours to allow the acidic conditions to produce NO gas. Crystals were harvested and flash frozen in 3.5 M sodium malonate pH 5 over 2-8 hours after sodium nitrite was added to the reservoir solution.

7.2.3. Serial Crystallographic Data collection

Serial datasets were collected from six nitrite soaked crystals and two NO soaked crystals to give a total of 159 datasets. Datasets for nitrite soaked crystals were collected at 100 K on Diamond Light Source beamline I04 using a Pilatus 6M detector. A total angular range of 38 degrees was collected with 0.1° oscillations and 0.05 s exposures per image at λ = 0.97 Å wavelength with a 90 µm x 45 µm beam and 20% transmission. This gave a total exposure time of 19 seconds and an average dose of 0.69 MGy per dataset as calculated by RADDOSE-3D (Zeldin et al., 2013) based on a

measured flux of $2e^{12}$ ph/s. NO soaked serial datasets were collected on Diamond Light Source beamline I02 using a Pilatus 6M detector over a total angular range of 30° at λ = 0.9795 Å wavelength with an 82 µm by 21 µm beam. Crystals had a total exposure of 30 seconds and average dose of 0.03 MGy per datasets as calculated by RADDOSE-3D (Zeldin et al., 2013) based on a measured flux of $2e^{10}$ ph/s. Serial datasets were collected of the same 38/30 degree regions of each nitrite and NO soaked crystal, respectively. Initial analysis of the starting structures from the eight crystals used was performed to determine the 'best' (i.e. highest occupancy) ligand-bound form. A total of 45 datasets were then selected from the 'best' nitrite soaked crystal along with 5 datasets from the 'best' NO soaked crystal, allowing for a detailed analysis of the proposed in-situ X-ray induced reaction mechanism of *Ac*NiR.

7.2.4. Data Processing

All datasets were integrated and scaled with XDS (Kabsch, 2010b) and merged with AIMLESS (Evans and Murshudov, 2013) to individual maximum high resolution limits based on mean $I/\sigma(I) > 2$ in the highest resolution bin. Similar scripts to those used in Chapters 2 and 3 were used to deal with the high volume of data to be processed. The structure of ligand-bound AcNiR was solved to atomic resolution by Antonyuk et al. (2005) (PDB code 2BWI) and here dataset 1 of the nitrite and NO soaked crystals were refined against 2BWI, with the nitrite removed, using REFMAC5 (Murshudov et al., 2011). Model building was carried out using COOT (Emsley et al., 2010). Subsequent datasets were refined against the final model of the previous dataset, i.e. the final model of dataset 8 was used as the starting model for dataset 9 etc. Nitrite, NO and water were modelled into the T2Cu active site F_o-F_c difference density. Model validation was performed using Molprobity (Chen et al., 2010). Figures of all structures were prepared with PyMol version 1.3r1 (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC).

7.3. Results and Discussion

7.3.1. Protein Production, Purification and Crystallisation

Production and purification of *Ac*NiR has been refined through previous studies giving up to 90 mg/ml of protein from 2 litres of LB culture. Samples of the supernatant were run on SDS-PAGE at various stages of ammonium sulphate precipitation, shown in Figure 7.4. Ammonium sulphate was gradually increased in samples shown between lanes 2 and 6 in Figure 7.4 until all *Ac*NiR precipitated (Fig. 7.4, lane 7), when the AcNiR was resuspended in 50 mM MES pH 6.5 (Fig. 7.4, lane 9). Although not 100% pure AcNiR is easily crystallised, with initial crystal trials producing small crystals (Fig. 7.5a). These were used as seeds to produce larger crystals over a few days at 19°C for high resolution data collection (Fig. 7.5b).

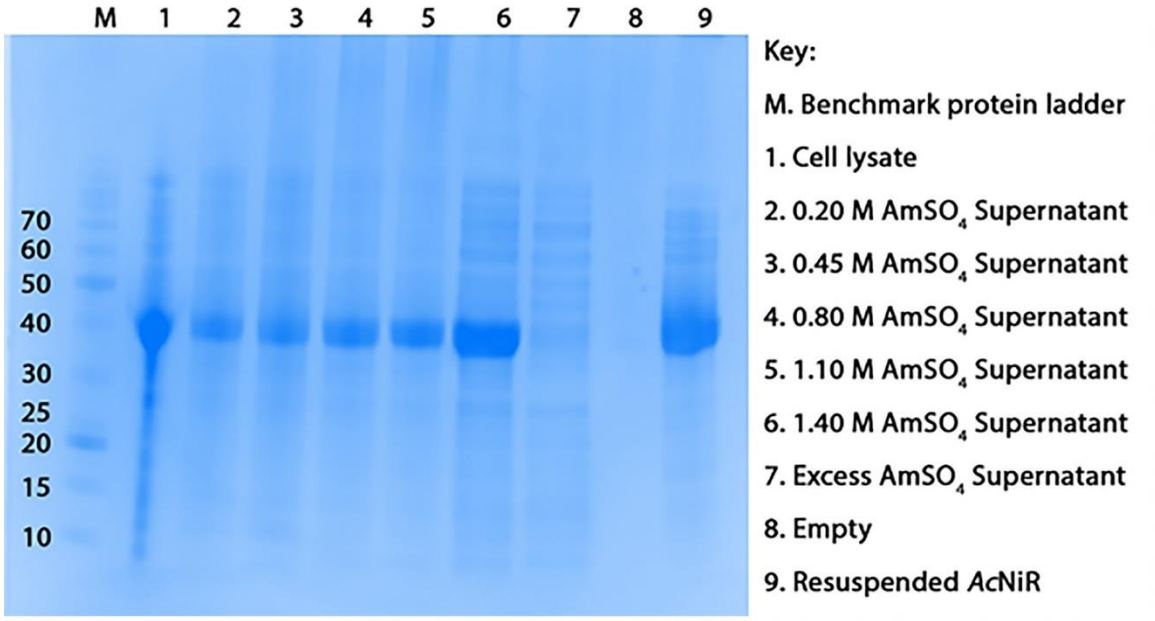


Fig. 7.4: SDS PAGE of ammonium sulphate precipitation purification of recombinant *Ac*NiR.

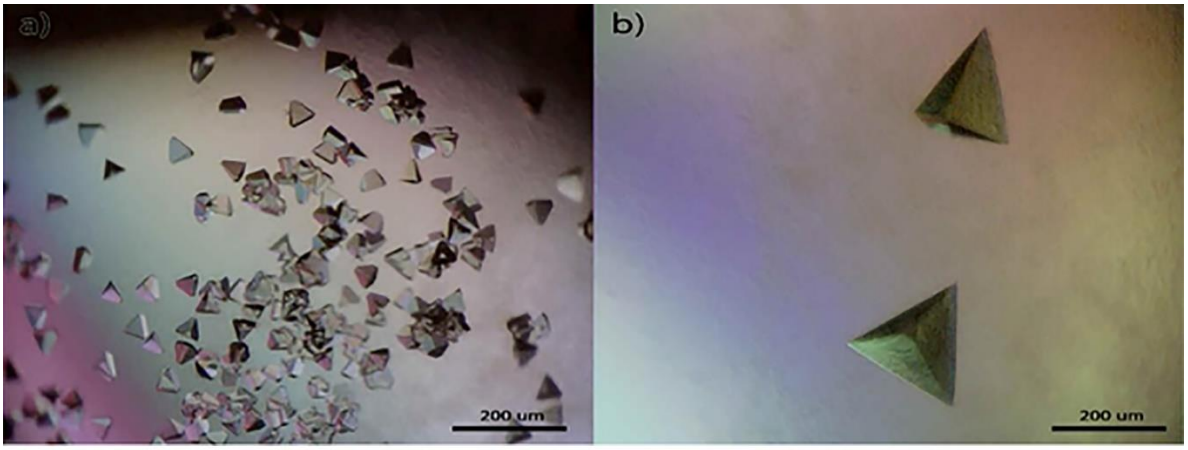


Fig. 7.5: *Ac*NiR crystals grown in ammonium sulphate, sodium acetate and MES with initial small crystals (a) and large crystals following seeding (b).

7.3.2. Serial Data Collection Statistics

Repeated exposure of the same region of the nitrite soaked crystal naturally caused a steady reduction in diffraction resolution and an increase in overall B-factors and unit cell parameters with increasing dose, as shown in figure 7.6a, b and c respectively. The high resolution limit decreased from atomic resolution, 1.07 Å initially, to 1.51 Å at 20 MGy exposure (the Henderson limit) (Henderson, 1990), and to 1.65 Å at the $D_{1/2}$, or Garman limit (Owen et al., 2006b), where the resolution drops to half that of the initial crystal resolution. This occurred after an average dose of 27.6 MGy (Fig. 7.6a). High resolution limit shows an unexpected decrease between datasets 7 and 10 (Fig. 7.6a) before returning to a value in step with initial data collections which continues to gradually decrease with increasing dose. As the high resolution returns to around 1.1 Å following this anomaly it is likely a result of data being cut unnecessarily during semi-automated processing. A similar event is seen in B-factors between datasets 7 and 10 with an unexpected rise which falls back to initial levels (Fig. 7.6b). Unit cell parameters remain fairly constant throughout the repeated exposures, varying by approximately 0.5 Å, including the unusually high first dataset, and approximately 0.4 Å if this dataset is considered as an outlier (Fig. 7.6c). The initial high resolution data allowed the position of nitrite to be modelled with relatively high certainty, but modelling becomes more difficult with decreasing resolution around the Henderson limit, and the

electron density is very difficult to interpret at the Garman limit, as discussed later in this chapter. Even with the initial atomic resolution data there are alternative interpretations that can be modelled based on electron density alone.

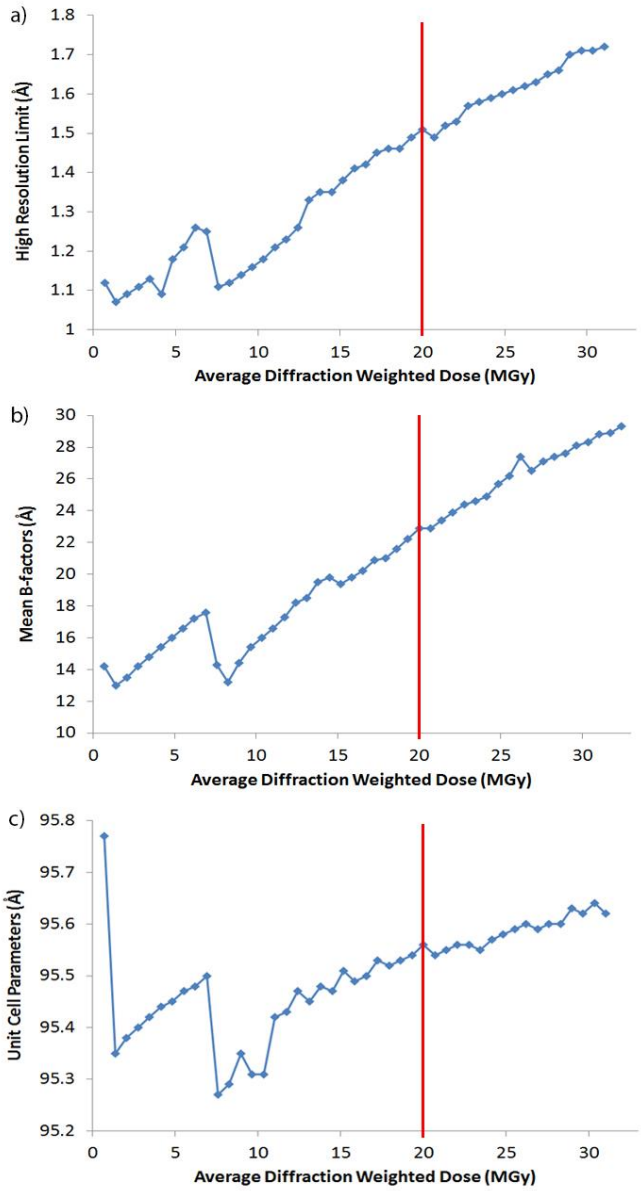


Fig. 7.6: Data quality graphs for serial data collection of 45 datasets from a single AcNiR crystal including high resolution limit based on an I/σ(I) > 2 (a), the mean B-factors of final models (b) and the unit cell parameters (c) plotted against average diffraction weighted dose as calculated by RADDOSE-3D (Zeldin et al., 2013). Red lines represent the Henderson limit at 20 MGy (Henderson, 1990).

7.3.3. Ligand Modelling

7.3.3.1. T2Cu NO_2 Modelling

Even at the high starting resolution of 1.07 Å the difference electron density at the location of ligand binding at the T2Cu site can be interpreted in various ways (Fig. 7.A1). Initial modelling of a single high occupancy nitrite molecule (70%) with a lower occupancy water molecule (30%) to give a fully occupied T2Cu site presents with no positive or negative difference density (Fig. 7.7a-b). However, B-factors for the central N-atom are approximately 5-6 $Å^2$ higher than those of the flanking oxygen atoms. The conformation of bound nitrite is thought to be variable from earlier studies (Antonyuk et al., 2005) and, consistent with this, the ligand difference density can be modelled using an oscillating/rotating ligand model with two nitrite molecules at 40% occupancy each (NO_2a and NO_2b) and a water molecule with 20% occupancy (Fig. 7.7c-d). Again, this model produces no significant positive or negative difference electron density and B-factors are more consistent for all the atoms of the nitrite molecules (~ 3-4 $Å^2$ difference between N and O atom values) compared to a single higher occupancy nitrite molecule. This represents a choice between two models: a simple one nitrite bound model and a more complex one with two alternate bound nitrite conformations.

The irregular shaped electron density was modelled above in both cases with a water molecule included to account for partial difference density. However, the density can also be modelled using a third nitrite molecule (NO_2c) with the O1 and N atoms interacting with the T2Cu and O2 pointing away from the T2Cu. Modelling three nitrite molecules (33% occupancy each) produces no positive or negative difference density and stable B-factors through all nitrite molecules. However, the conformation of NO_2c and its interactions with the T2Cu atom make little chemical sense so is unlikely to be even a partial or transient true state of the active site. Based on these observations, the first three datasets measured on the crystal were modelled with two nitrite molecules (NO_2a and NO_2b) to represent the oscillating/rotating N atom of the nitrite ligand (Fig. 7.7c-d). The water molecule seen in dataset one is not observed in datasets two and three. By

dataset four the nitrite molecule has stabilised into the “side-on” symmetrical orientation, labelled NO_2b in this study, which was previously observed by Antonyuk et al. (2005). From this point a single nitrite molecule in the NO_2b conformation at an occupancy of 60% is modelled up to dataset ten (Fig. 7.10d).

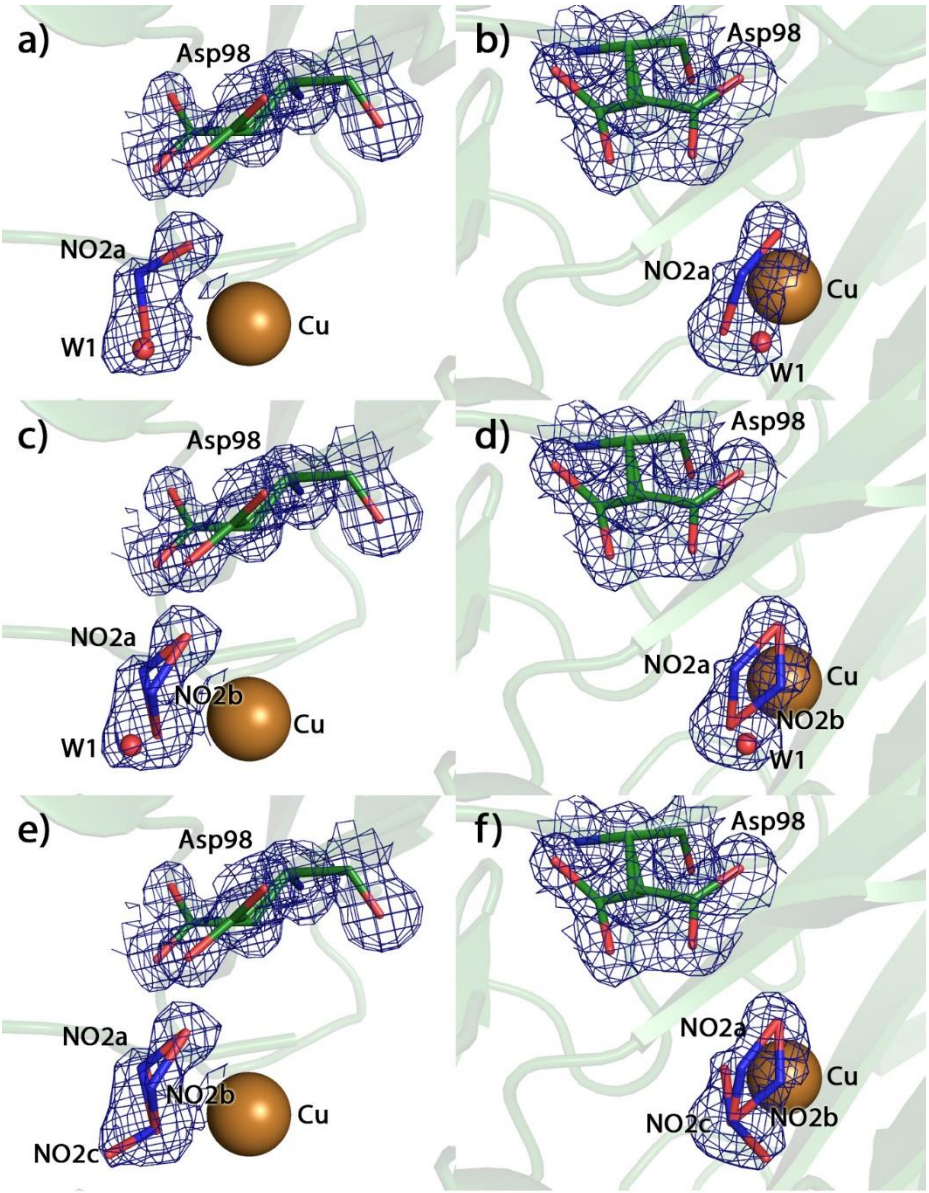


Fig. 7.7: Nitrite modelling options for dataset one of the nitrite soaked serial data collection with side on and overhead view of $2F_o$-F_c electron density at σ = 1. All Asp98 residues at modelled at 50% occupancy for gatekeeper and proximal conformations. a) Modelling of a single nitrite molecule (NO2a) at 70% occupancy with a water molecule (W1) at 30% occupancy. b) Dual occupancy nitrite (NO2a/NO2b) at 40% occupancy each with a water molecule (W1) at 20% occupancy. c) Modelling of three nitrite molecules at 33% occupancy each. Asp98 and nitrite are represented as sticks and water molecules and copper atoms as spheres. All atoms are coloured by element.

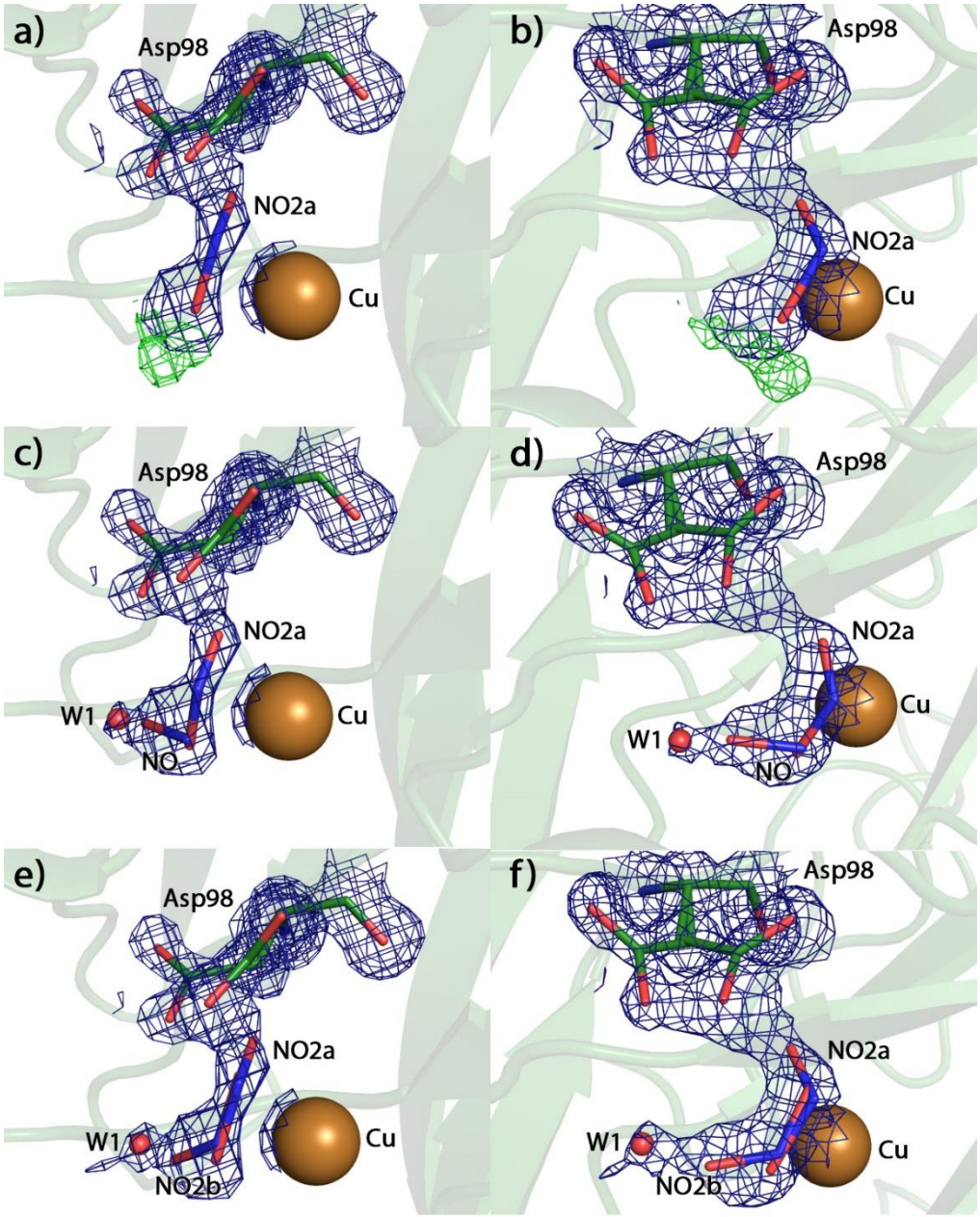


Fig. 7.8: Nitrite modelling options for dataset one eleven of the nitrite soaked serial data collection with side on and overhead view with $2F_o$-F_c electron density at σ = 1 in blue and F_o-F_c density at σ = 3 in green. All Asp98 residues at modelled at 40% occupancy for gatekeeper and 60% occupancy for proximal conformations. a) A single side-on nitrite molecule at 70% occupancy. b) A single side on nitrite molecule at 70% occupancy with two waters W1 and W2 at 30% and 80% occupancy, respectively. c) Dual occupancy nitrite and NO molecules at 33% occupancy each with two waters (W1 and W2) at 33% and 80% occupancy, respectively. d) Dual occupancy nitrite and NO molecules at 33% occupancy each with NO in an end on orientation leaving the active site and a water molecule at 80% occupancy. e) Dual occupancy nitrite and NO at 33% occupancy with a second NO (NOb) leaving the active site at 35% occupancy. f) Dual occupancy NO modelled as the forth ligand to the T2Cu (NOa) at 35% occupancy and a second NO leaving the active site (NOb) at 35% occupancy. g) Dual occupancy NO modelled as the forth ligand to the T2Cu (NOa) at 35% occupancy and a second NO leaving the active site (NOb) at 35% occupancy with a water molecule (W2) at 80% occupancy. Asp98, nitrite and NO are represented as sticks and water molecules and copper atoms as spheres. All atoms are coloured by element.

By dataset eleven the nitrite soaked crystal has been exposed to a 7.59 MGy dose of X-ray radiation and is still diffracting to around 1.1 Å resolution (Fig. 7.6a) when new positive difference density is observed in the active site suggesting a previously unseen active site conformation (Fig. 7.8a). As with dataset one the active site density has several possible interpretations (Fig.7.8). This density can be modelled as a side on nitrite molecule with two water molecules (Fig. 7.8b), a lower occupancy side on nitrite molecule and NO molecule at the T2Cu with two water molecules (Fig. 7.8c), a side on nitrite molecule with an NO molecules leaving the T2Cu site with a water (Fig. 7.8d), a side on nitrite and NO molecule at the T2Cu site and a second NO molecule leaving the T2Cu site (Fig. 7.8e), one NO molecule at the T2Cu site and one leaving the T2Cu site (Fig. 7.8f) and finally one NO molecule at the T2Cu site, one leaving the T2Cu site and a water molecule (Fig. 7.8g).

The characteristic curved shape of the density by dataset eleven, apparent in all side on views in Figure 7.8 (left hand side images), suggests the nitrite remains in the active site, so on the basis of the structural interpretation alone the NO-NO (Fig. 7.8f) and NO-NO-W (Fig. 7.8g) models are unlikely. The split occupancy of Asp98 proximal and gatekeeper conformations is also not characteristic of a solely NO occupied T2Cu (NO-NO and NO-NO-W); or a completely nitrite occupied T2Cu site (NO_2-W-W (Fig. 7.8b)) (Antonyuk et al., 2005). This leaves the NO_2-NO-W-W (Fig. 7.8c), NO_2-NO-W (Fig. 7.8d) and NO_2-NO-NO (Fig. 7.8e) T2Cu site models to be considered. The presence of NO bound to the T2Cu in a bidentate manner in later datasets suggests NO is not leaving the T2Cu at this stage so the NO2-NO-W (Fig. 7.8d) and NO_2-NO-NO (Fig. 7.8e) models can be discounted. This leaves the NO_2-NO-W-W model (Fig. 7.8c) which makes chemical sense and could be justified based on the observed Asp98 conformation. The following “catalytic movie” was generated with this NO_2-NO-W-W model.

7.3.4. NO_2 Soak Movie

Over the course of the forty-five datasets collected from the same nitrite soaked crystal the T2Cu active site goes through several conformational changes that can be observed thanks to

the high resolution diffraction data, even if difficulties remain in interpretation. Over the first three datasets nitrite appears to oscillate/rotate horizontally with the O1 and O2 atoms fixed in place and the N atom moving between two positions labelled NO_2a and NO_2b in Figure 7.10a-c. This explains the higher B-factors observed of the N atom when a single high occupancy nitrite molecule was modelled into the difference density. In dataset one the electron density covers all atoms in both the proximal and gatekeeper conformations of Asp98, with both positions modelled at 50% occupancy (Fig. 7.10a). The density decreases slightly in the gatekeeper conformation for the Oδ1, Oδ2 and Cγ atoms over datasets two to four (Fig. 7.10b-c) so was modelled with 40% occupancy, with 60% for the proximal conformation. At this stage nitrite has stabilised as NO_2b (Fig. 7.10d), referred to as the "side on" conformation by Antonyuk et al. (2005). Nitrite remains in this position for a further six datasets where it is optimally positioned for His255 to donate the second proton for cleavage of the O-NO bond. Between dataset one and dataset four the distance between His255 Nε2 and T2Cu decreased by 0.08 Å from 4.17 Å to 4.09 Å when nitrite has stabilised in the NO_2b conformation. At this point the crystal has been exposed to a total average dose of 2.76 MGy (13.8 % of the Henderson limit) and the resolution has remained constant around 1.1 Å (Fig. 7.6a).

At dataset eleven a previously unseen T2Cu-ligand conformation is observed with a new water appearing, labelled WPc in Figure 7.10e. Based on the position of WPc at the T2Cu catalytic centre and the dual occupancy of Asp98 in the proximal and gatekeeper positions this has been modelled as an intermediate stage where both nitrite, NO and a water are present in the crystal at 30% occupancy each to give a total occupancy of 90% for the T2Cu active site (Fig. 7.10a). At this stage the crystal has been exposed to a total average dose of 7.59 MGy and the high resolution limit is still unchanged around 1.1 Å. W2 in Figure 7.10 appears and disappears over the course of the next five datasets (Fig. 7.10e-i). Based on the split occupancy of substrate and product at the T2Cu active site this water may represent the final water in the water chain between the T1Cu and T2Cu to allow proton transfer and reduction of the T2Cu site to facilitate the second protonation and cleavage of nitrite. The water chain will be discussed in more detail later in this chapter. By dataset

sixteen the crystal has been exposed to 11.04 MGy (55.2 % of the Henderson Limit), the resolution has decreased to 1.21 Å and the occupancy of Asp98 favours the proximal conformation at 60% (Fig. 7.9b).

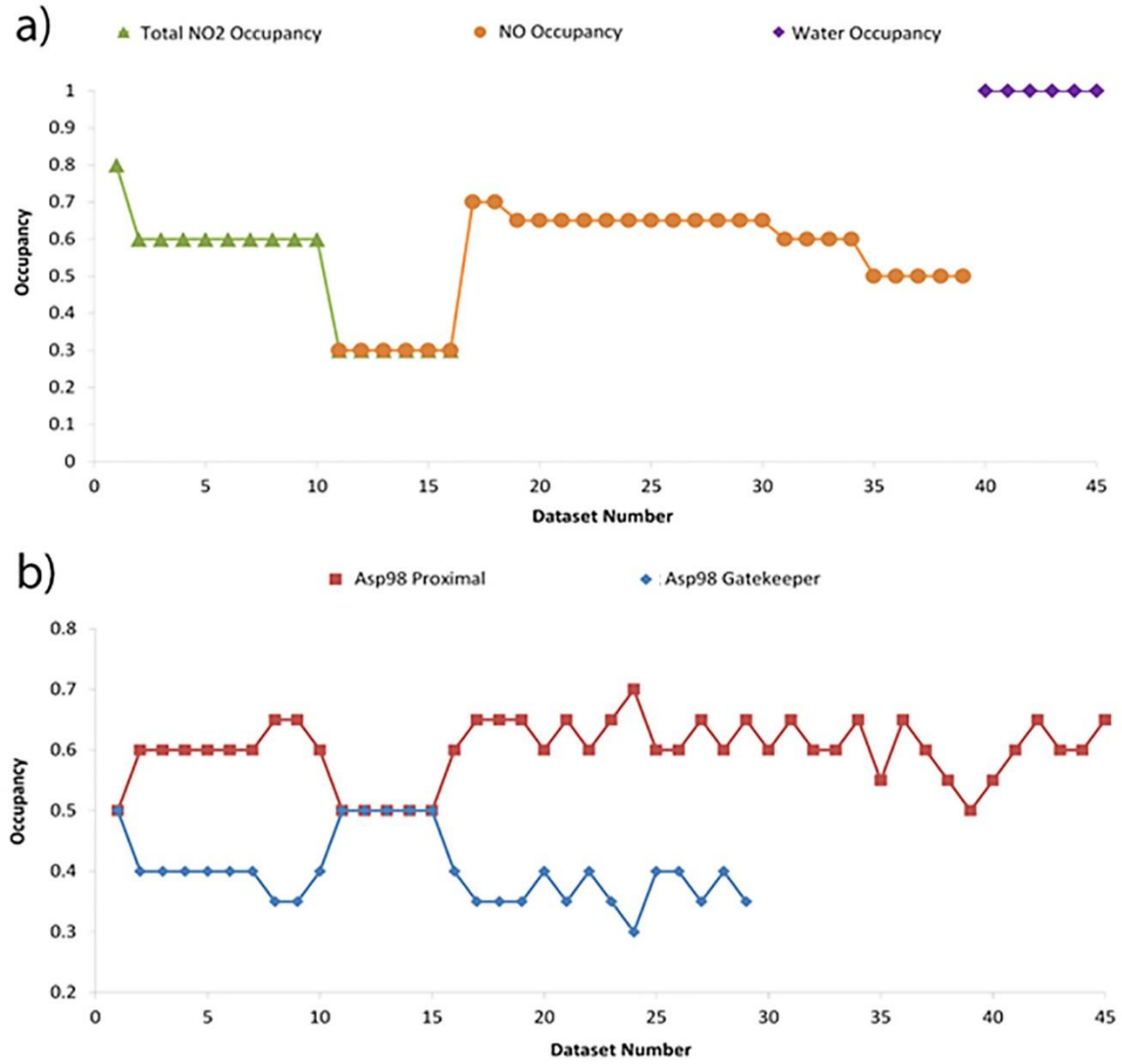


Fig. 7.9: Ligand and Asp98 conformation occupancy plotted against increasing dataset number. a) Ligand occupancy with total nitrite occupancy (the sum of multiple conformations when present) as green triangles, NO occupancy as orange circles and water occupancy as purple diamonds. b) Asp98 alternate conformation occupancy with the proximal conformation as red squares and the gatekeeper conformation as blue diamonds.

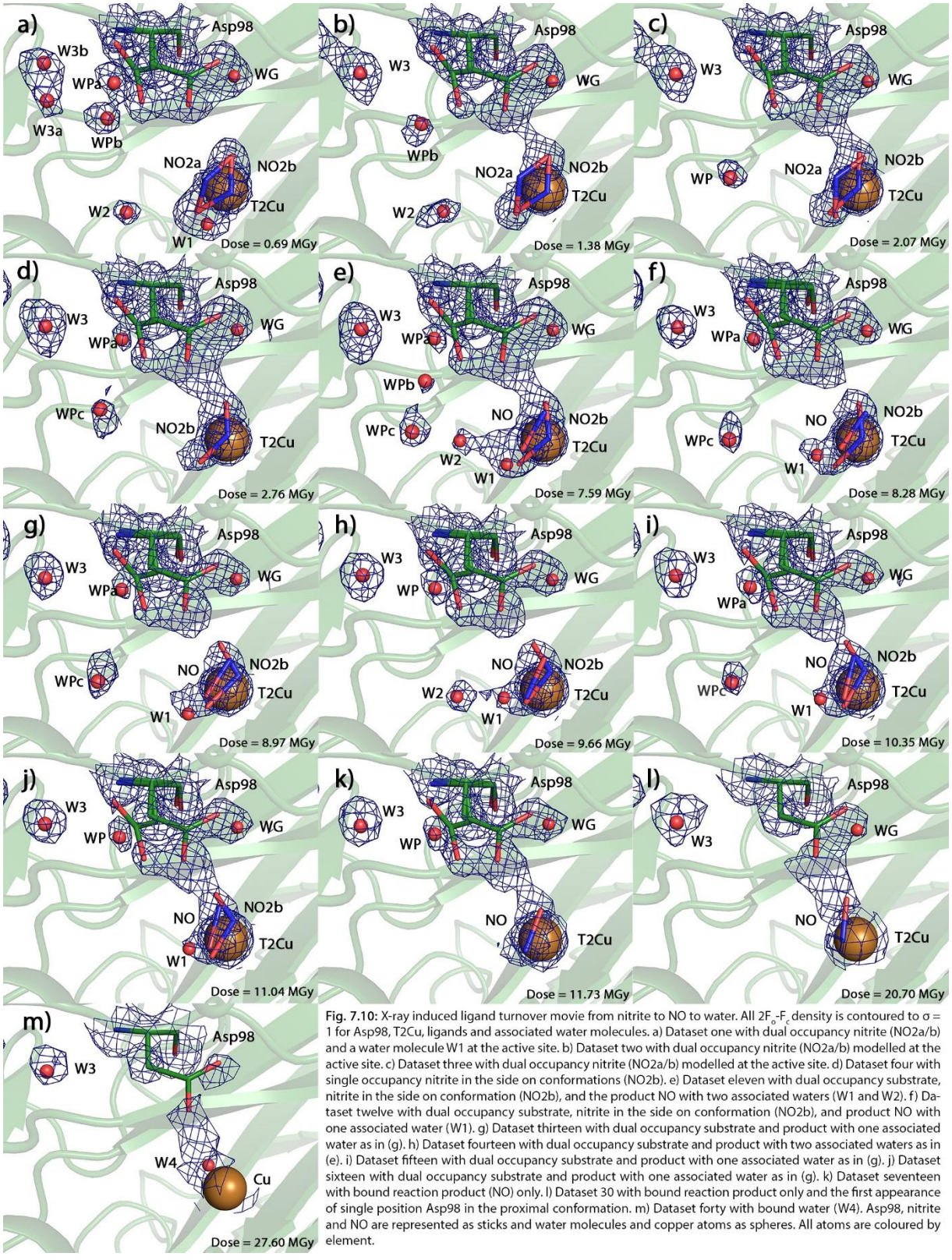


Fig. 7.10: X-ray induced ligand turnover movie from nitrite to NO to water. All $2F_o$-F_c density is contoured to σ = 1 for Asp98, T2Cu, ligands and associated water molecules. a) Dataset one with dual occupancy nitrite (NO2a/b) and a water molecule W1 at the active site. b) Dataset two with dual occupancy nitrite (NO2a/b) modelled at the active site. c) Dataset three with dual occupancy nitrite (NO2a/b) modelled at the active site. d) Dataset four with single occupancy nitrite in the side on conformations (NO2b). e) Dataset eleven with dual occupancy substrate, nitrite in the side on conformation (NO2b), and the product NO with two associated waters (W1 and W2). f) Dataset twelve with dual occupancy substrate, nitrite in the side on conformation (NO2b), and product NO with one associated water (W1). g) Dataset thirteen with dual occupancy substrate and product with one associated water as in (g). h) Dataset fourteen with dual occupancy substrate and product with two associated waters as in (e). i) Dataset fifteen with dual occupancy substrate and product with one associated water as in (g). j) Dataset sixteen with dual occupancy substrate and product with one associated water as in (g). k) Dataset seventeen with bound reaction product (NO) only. l) Dataset 30 with bound reaction product only and the first appearance of single position Asp98 in the proximal conformation. m) Dataset forty with bound water (W4). Asp98, nitrite and NO are represented as sticks and water molecules and copper atoms as spheres. All atoms are coloured by element.

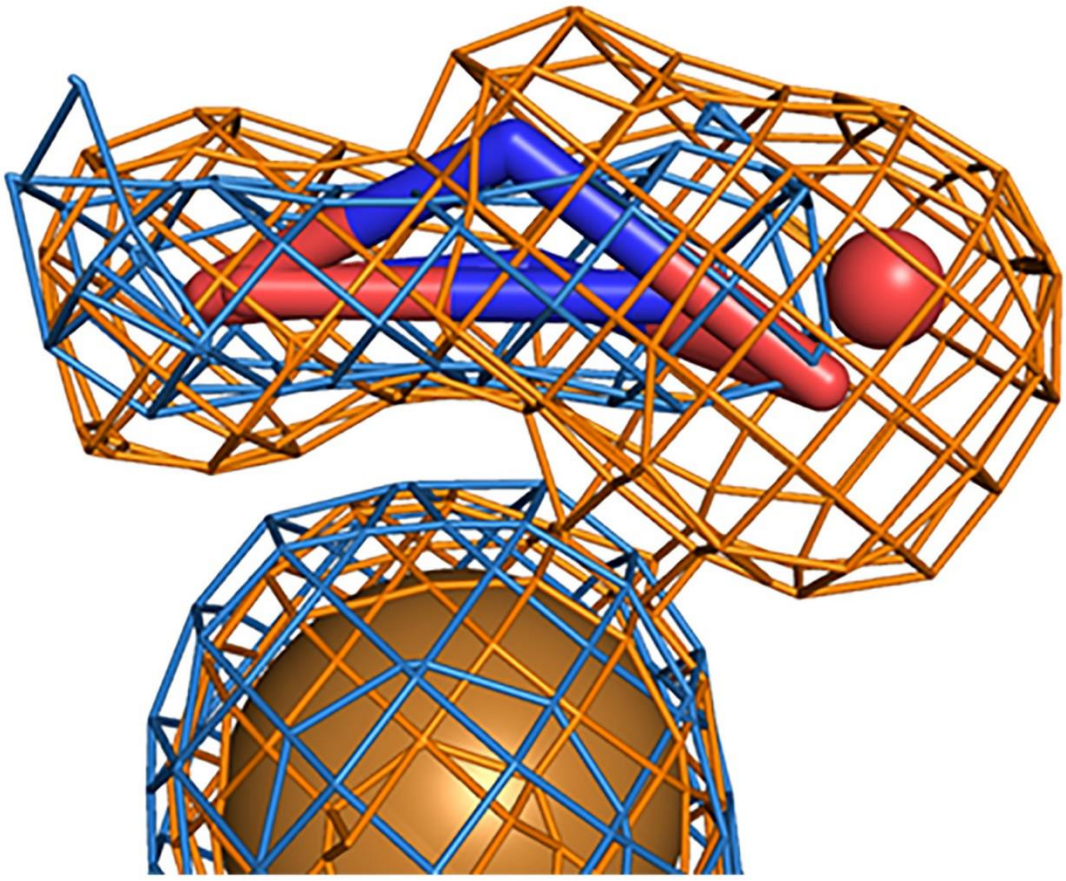


Fig. 7.11: Close up of the T2Cu site with nitrite bound dataset one electron density in orange at 1.09 Å resolution superimposed against NO bound dataset seventeen electron density in blue at 1.21 Å resolution. Both densities are contoured to σ = 1. Nitrite and NO are represented as sticks and water molecules and copper atoms as spheres. All atoms are coloured by element.

By dataset seventeen the occupancy of Asp98 in the proximal position has reached 65%, NO occupancy has been modelled in line with this occupancy in the active site (Fig. 7.9). The curved density shape typical of a nitrite bound T2Cu site, coloured in orange in figure 7.11, has become more flat relative to the copper atom, shown as blue density in Figure 7.11. At this stage the crystal has been exposed to an average dose of 11.73 MGy and the high resolution limit is 1.23 Å. Previously the mode of binding of the enzymatically generated NO product at the T2Cu centre has been met with lively discussion. Spectroscopic data suggests an end-on mode of binding while the first NO soaked structure of a CuNiR from *A. facaelis* (Tocheva et al., 2004) demonstrated a side-on mode. The non-enzymatic origin of the NO and the discovery that exposure to excess NO gas can cause a reversible reaction, as demonstrated by the production of red coloured NiR species (Ghosh et al.,

2007), complicated the analysis further. Endogenous side-on T2Cu-NO was also reported for *Ac*NiR by Antonyuk et al. (2005), however it was unclear if the presence of NO was the result of *in vivo* catalysis or X-ray induced turnover (Hough et al., 2008b). The structures presented here provide powerful evidence to resolve this issue with initial soaked nitrite being converted to a stable side on NO molecule in the crystal with near equidistant Cu-N and Cu-O bonds.

Following dataset seventeen there is no significant change in the appearance of the electron density until dataset thirty (Fig. 7.10l) where all density for the gatekeeper conformation of Asp98 has gone and NO moves closer to Asp98. However, at this stage in the data collection electron density at the active site is becoming very weak after being exposed to a total average dose of 20.70 MGy (103.5% of the Henderson Limit) and the high resolution limit decreasing to 1.49 Å. Electron density above the T2Cu decreases further over the remaining fifteen datasets and modelling becomes increasingly uncertain. A water molecule was modelled in dataset forty but by this stage the initial resolution has almost halved to 1.65 Å and the crystal had been exposed to a dose of 138% of the Henderson limit. To better investigate the T2Cu site serial data collection was attempted on a NO soaked crystal. Data processing and refinement statistics for all NO_2 soak movie structures are presented in Tables 5.A1-3.

7.3.5. NO-soaked crystals

The idea was to obtain higher resolution NO-bound structures than those resulting from the radiolysis experiment. However, the NO soaked crystals suffered considerably during the soaking and harvesting processes, with the best crystal obtained reaching a high resolution limit of only 1.76 Å. NO soaking was performed by mixing sodium nitrite with acidic conditions to cause the unstable nitrite to breakdown and produce NO gas in the sealed well. Crystals soaked for longer periods of time (24 hours) also showed a significant drop in high resolution diffraction suggesting the NO gas had a detrimental effect on the crystals. As with the nitrite modelling, there were multiple interpretations of the ligand density at the T2Cu site, exacerbated by the relatively low resolution

data. A single 100% occupancy NO molecule fitted into the electron density map at the T2Cu site resulted in positive difference density at 3σ (coloured green in Figure 7.12a) at either end of the N and O atoms, which suggests additional molecules, or a single longer molecule are present at the active site. Modelling an NO molecule with a water molecule at 65% and 35% occupancy, respectively, produces some positive difference density and exceedingly low B-factors for the water molecule (6.86 Å), almost two times lower than the T2Cu. Modelling both the NO and water molecule with an occupancy of 90% each, representing the products of nitrite cleavage, shows no difference density (Fig. 7.12b). However, both NO bonds are occupied by the T2Cu atom so the water has no ligands and could only be present as an alternate conformation not a liganding molecule to NO and the T2Cu. The density is large enough and has the three dimensional shape to accommodate two NO molecules, however, two NO molecules at 50% occupancy each resulted in positive difference density at 3σ (Fig. 7.12c). Finally, a nitrite molecule was modelled in the density (Fig. 7.12d) at 75% occupancy. Based on the length of the density and curved shape around the copper, although the resolution is not high enough to show a distinct separation from the copper atom electron density - like that observed in Figure 7.11 – bound nitrite is the most likely explanation. Thus, it appears that NO soaking was unsuccessful and nitrite has bound to the T2Cu instead. It is likely that some nitrite from the reservoir solution was no broken down to NO by the acidic conditions and diffused into the crystals. As a result the NO soaked data was not used further for the movie. Due to time constraints on this thesis the NO-soaking experiment was not repeated. However, this is a clear objective for future work.

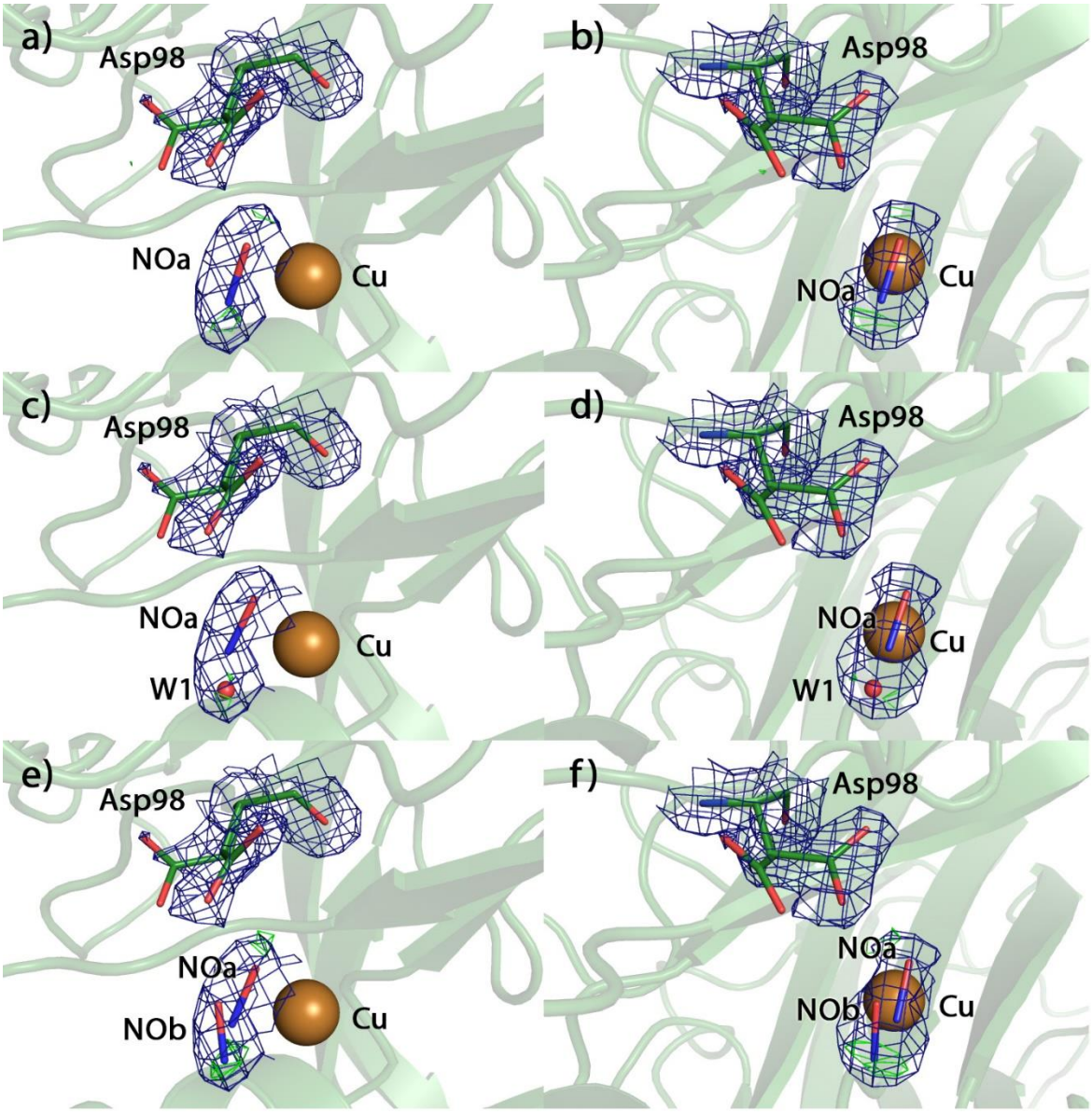


Fig. 7.12: Modelling options for NO soaked datasets with $2F_o$-F_c density contoured to σ = 1 and F_o-F_c density contoured to σ = 3 for Asp98 and all ligands at 1.76 Å resolution. a) Dataset one with 100% occupancy NO ligand. b) Dataset one with 100% occupancy NO and water. c) Dataset one with dual occupancy NO molecules (NOa/b) bound at 50% occupancy each. Dataset one with 75% occupancy nitrite bound. Asp98, nitrite and NO are represented as sticks and water molecules and copper atoms as spheres.

7.3.6. The Role of Asp98 in Ligand Sensing

Over the course of the in-crystallo catalytic reaction the occupancy of Asp98 in the gatekeeper and proximal conformations varies in line with the presence of the ligand binding in the T2Cu active site, be it either dual occupancy water/nitrite, nitrite, dual occupancy nitrite/NO or NO (Fig. 7.9). As observed in previous nitrite bound structures (PDB code: 2BWI), Asp98 shows equal

dual conformations of the gatekeeper and proximal positions, likely due to the incomplete incorporation of nitrite as evident by the presence of a water molecule at the T2Cu (Fig. 7.10a). By datasets two and three the water molecule has left, leaving dual conformation nitrite molecules and 60% occupancy of proximal Asp98, in line with the decreasing $2F_o$-F_c density at σ = 1 for Asp98-Oδ2 and –Cγ in the gatekeeper conformation (Fig. 7.10b-c). This occupancy remains in line with NO_2 occupancy up to dataset eight when Asp98-Oδ1, -Oδ2 and –Cγ $2F_o$-F_c density at σ = 1 decreases for the gatekeeper conformation and occupancy is reduced to 35% (Fig. 7.9b). By dataset eleven $2F_o$-F_c density at σ = 1 has returned to similar levels observed in dataset one and occupancy has returned to an even split between gatekeeper and proximal conformations (Fig. 7.9b). This is attributed to the dual occupancy of nitrite and NO (Fig. 7.9a) in the active site and persists up to dataset fifteen when $2F_o$-F_c density at σ = 1 for Asp98-Oδ1 in the gatekeeper conformations begins decreasing again (Fig. 7.10i). The occupancy of the gatekeeper conformation of Asp98 was reduced to 40% occupancy in dataset sixteen and further to 35% in dataset seventeen when NO is presented with 70% occupancy. Following dataset seventeen the occupancy of NO gradually reduces from 70% to 50% by dataset thirty nine when water is modelled in thereafter (Fig. 7.9a). Over this period Asp98 occupancy varies between 50-65% occupancy for the proximal conformation with all density lost for the gatekeeper conformation following dataset twenty nine (Fig. 7.9b). However, it is unclear if this is attributed to the presence of NO in the active site or radiation damage preventing accurate modelling of the residue.

Over the course of thirty datasets, when the crystal reached the Henderson dose limit, the proximal conformation of Asp98 correlates with the presence of different ligands at the T2Cu. Occupancy of the proximal conformation increases as nitrite stabilises in the side on conformation then drops with the dual occupancy of nitrite and NO only to rise again as nitrite is lost leaving T2Cu-NO in the active site. This suggests that Asp98 plays a role in sensing/guiding/stabilising the T2Cu ligand in place. The single conformation of Asp98 in the proximal positions following dataset twenty nine may be a true conformational change as it has been observed in high resolution NO soaked

*Ac*NiR structures before (PDB code: 2BW4). However, the presence of dual conformations of Asp98 in previously published native *Ac*NiR structures (PDG code: 2BW4) suggests the loss of the Asp98 gatekeeper conformations in dataset 40 onward in serial data collections is likely the result of radiation damage and not a true conformational change. In the following section T1Cu and T2Cu active sites have been compared with previously published data to further investigate these changes in the active site.

7.3.6.1. Comparison with Published Ligand Positions

Atomic resolution structures of nitrite bound, NO bound and native states (PDB codes: 2BWI, 2BW5 and 2BW4, respectively) by Antonyuk et al. (2005) were used for comparison with complementary stages of the reaction as observed through the serial data collection. The overall fold of the protein is unchanged. Comparison of the T1Cu site between published data and serial data collected from nitrite and NO bound crystals shows little to no variation between the positions of the metal coordinating His95, Cys136, His145 or Met150 at all stages of the reaction (Fig. 7.13a, c and e). All structures from serial data collection have a Cu-Cys-Sγ bond length of 2.2 Å, consistent with a reduced T1Cu site (Hough et al., 2008b, Ellis et al., 2002). T1Cu-ligand bond lengths for the structures used in the nitrite soak movie (Fig. 7.10) are listed in Table 7.A4-7.

Nitrite bound AcNiR shows little variation between the T2Cu liganding residues (His100, His135 and His306B), the secondary proton donor (His255B) or the ligand guiding Asp98 with the side chain in both the proximal and gatekeeper conformations in both structures (Fig. 7.13b) (Antonyuk et al., 2005). Nitrite in 2BWI (coloured in blue in Figure 7.13b) is modelled at 40% occupancy with the water molecule W1 accounting for the remaining occupancy as the fourth ligand to the T2Cu atom. Nitrite bound in dataset one from the serial data collections shows a much higher occupancy of bound nitrite with a total of 80% from the two alternate conformations and a water molecule at 20% occupancy. This may be the result of more effective uptake of nitrite through prolonged soaking times with higher occupancy producing this previously unseen dual conformation

of nitrite. Both 2BWI and dataset one were refined to around 1.1 Å so a dual occupancy should have been observed in 2BWI if it was present in the crystal. The nitrite conformation observed in 2BWI is similar to the NO_2b conformation observed in dataset one (Fig. 10a) with both binding the T2Cu in a bidentate manner with Cu-O distances of 1.98/2.19 Å, 2.07/2.05 Å and 2.06/2.04 Å for 2BWI nitrite, dataset one NO_2a and dataset one NO_2b, respectively.

The NO bound T2Cu active site from 2BW5 and dataset seventeen of the nitrite soak serial data collection again show little to no variation between the T2Cu liganding residues His100, His135 and His306B or the second proton donor His255B (Fig. 7.13d). Asp98 from 2BWI is in the proximal conformation with atomic resolution data allowing two proximal conformations to be modelled. Dataset seventeen on the other hand shows both gatekeeper and proximal conformations of Asp98, likely the result of incomplete occupancy of NO in the active site (Fig. 7.13d). However, Asp98 occupancy is in favour of the proximal conformation at 60% (Fig. 7.9a). NO molecules for 2BW5 and dataset seventeen are in similar side on orientations to T2Cu with N/O-T2Cu distances of 2.1 Å and 2.0 Å, respectively. Distances from Asp98 vary slightly more with 2WB5 at 2.7 Å and dataset seventeen at 2.4 Å. The similarity between ligand modelling of both nitrite and NO between soaked structures and serial data collections suggests fast data collections could be a valuable tool for tracking the catalytic processes carried out by metallo-proteins. However, the technique does suffer for gradual loss of resolution and radiation damage with increasing dose. This drawback is evident when modelling the final step of the reaction where NO dissociated and is replaced by a water molecule to returning the enzyme to its native state. This problem might be successfully addressed in the future by using faster X-ray detectors for more rapid data collection.

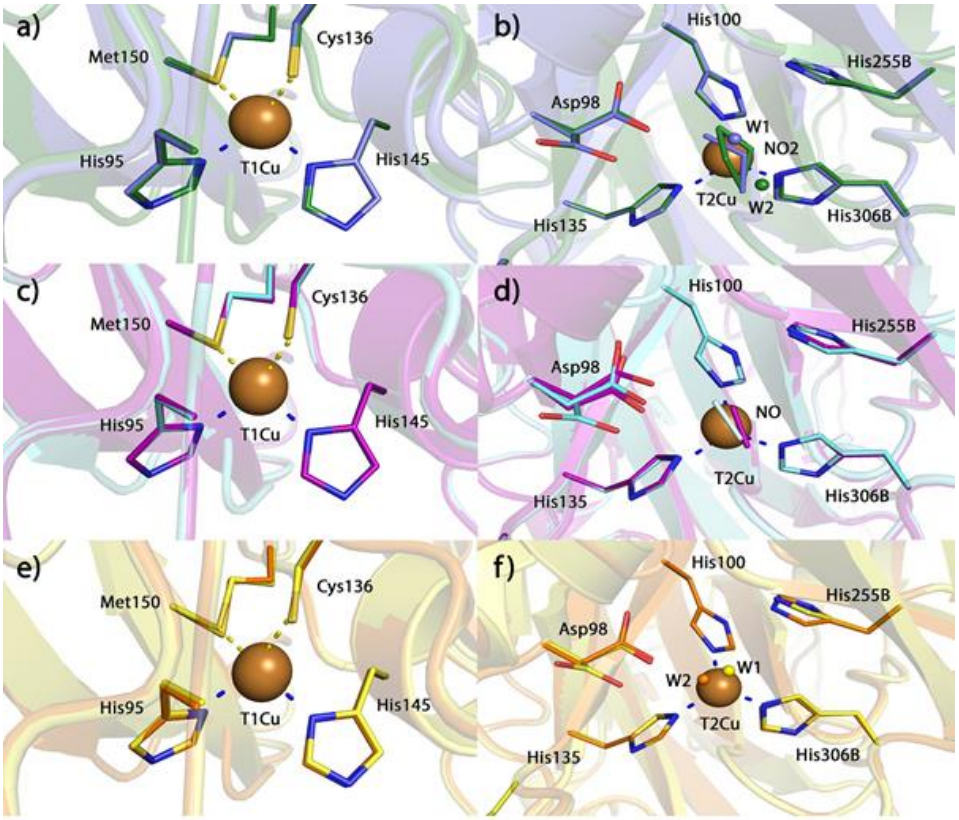


Fig. 7.13: Superposition of T1Cu and T2Cu sites of serial data collection models against published nitrite bound, NO bound and native structures. Residues labelled B represent residues from a second monomer which completes the T2Cu site. a) Dataset one (green) T1Cu site superimposed with 2BWI (blue). b) Dataset one (green) T2Cu site superimposed with 2BWI (blue). c) Dataset seventeen (cyan) T1Cu site superimposed with 2BW4 (purple). d) Dataset seventeen (cyan) T2Cu site superimposed with 2BW4 (purple). e) Dataset forty (orange) T1Cu site superimposed with 2BW5 (yellow). f) Dataset forty (yellow) T2Cu site superimposed with 2BW5 (orange). All residues, nitrite and NO are represented as sticks and water molecules and copper atoms as spheres.

The published high resolution native structure of *Ac*NiR shows a single water molecule 2.1 Å from the T2Cu and 2.8 Å from the proximal conformation of Asp98. The water modelled into serial dataset forty is closer to both the T2Cu and Asp98 at 1.9 Å and 2.1 Å, respectively (Fig. 7.13f). At this stage, dataset forty has reached 138% of the Henderson limit and the high resolution limit has decreased by almost half the initial high resolution limit to 1.65 Å. Although this is still high enough resolution to model well conserved waters with confidence, continued exposure has clearly decreased the quality of the electron density maps. To better understand how *Ac*NiR returns to the resting state, high resolution data from an NO soaked crystal is necessary as data quality has suffered to the point where modelling cannot be performed unambiguously and with confidence. Alternatively, data collection performed at higher cryogenic temperatures than the standard 100K

used here may allow a faster turnover of substrate with equivalent dose, permitting the full reaction to be observed over fewer data collections and less damage.

7.3.7. T1Cu-T2Cu Water Chain, a Possible Hydrogen Bonded Chain/ Proton Wire

A water chain lies on the interface between monomers of the trimer structure, Figure 7.14 shows the surface of monomer A (dataset one from the nitrite soaked series) and an approximately 30 Å long water channel connecting the T1Cu (buried under the surface) to the T2Cu site. Analysis of serial datasets of *Ac*NiR demonstrates a well conserved chain of approximately nine water molecules between the T1Cu and T2Cu (Fig. 7.15). W8Pa-c and WG in Figure 7.14 represent partial occupancy waters present with the proximal and gatekeeper conformations of Asp98, respectively. I.e. in unit cells with the proximal conformation, W8P occupies this position as the gatekeeper is not there to

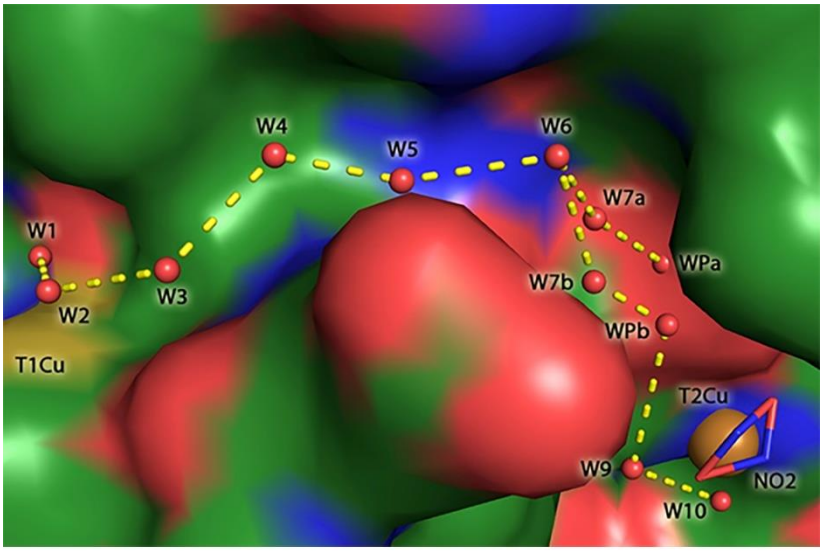


Fig. 7.14: The T1Cu-T2Cu water channel of dataset one of nitrite soaked AcNiR with water molecules and copper atoms as spheres, nitrite as sticks and the protein represented as a surface. Dashed yellow lines represent the potential proton transport pathway between the copper atoms.

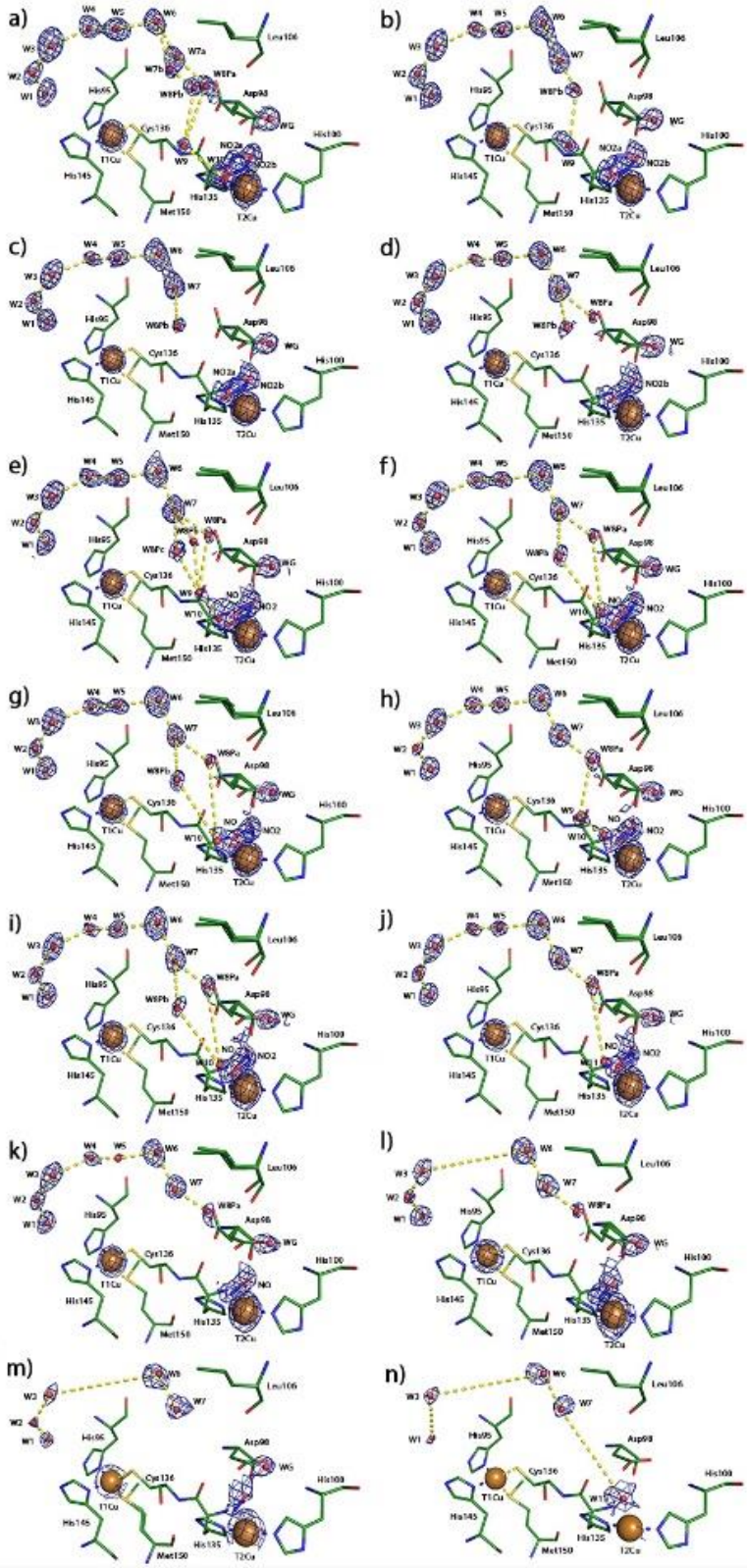
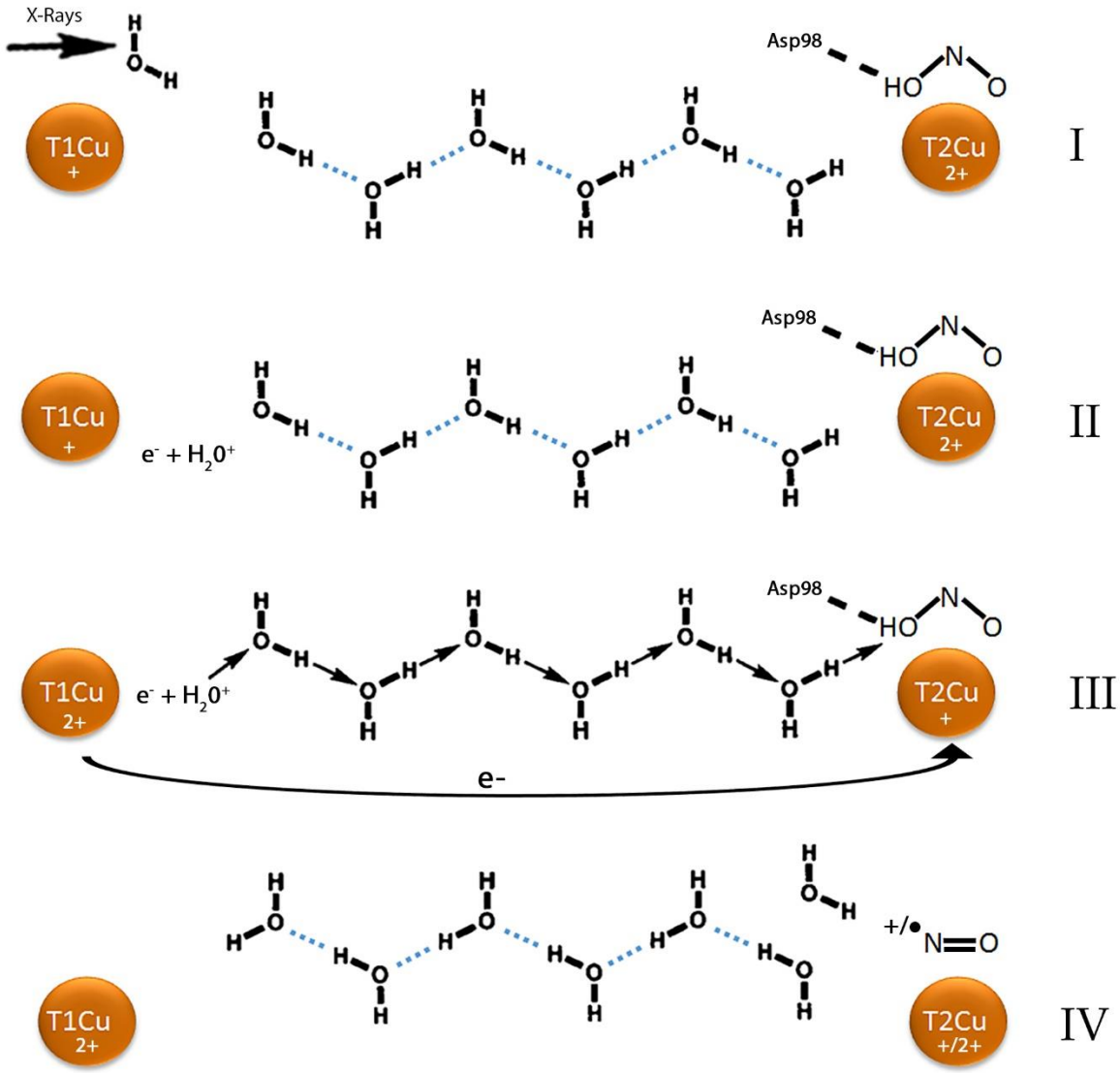


Fig. 7.15: A movie of the changes in the T1cu-T1Cu water chain as the reaction progresses at the T2Cu active site. All 2Fo-Fc density around ligands, water molecules and copper atoms is contoured to σ = 1. Dashed yellow lines represent potential electron transport pathways between water molecules. Waters labelled W8P and WG represent partial occupancy waters present when the proximal and gatekeeper conformations of Asp98, respectively, are observed. a) Dataset one. b) Dataset two. c) Dataset three. d) Dataset four. e) Dataset eleven. f) Dataset twelve. g) Dataset thirteen. h) Dataset fourteen. i) Dataset fifteen. j) Dataset sixteen. k) Dataset seventeen. l) Dataset twenty-three. m) Dataset thirty. n) Dataset forty. All residues, nitrite and NO are represented as sticks, waters and copper atoms as spheres and all coloured by element.

displace it. The water chain between W1-W7 is well conserved with waters remaining in fairly constant positions over repeated exposures. A double conformation of W7 is seen in dataset one (Fig. 7.15a) but is not observed in later datasets. After W7 the water chain becomes more variable, particularly W8P which can be seen in three different positions over the serial data collection. This is most notable in dataset eleven (Fig. 7.15e) where W8Pa, W8Pb and W8Pc are all seen and the W1-W9 chain to the T2Cu is complete. W10 represents a partial occupancy water as the fourth ligand of T2Cu where either both nitrite and NO have left the active site or were not present from the beginning so is not likely a part of the water chain.

The water chain may represent a “hydrogen bonded chain” or “proton wire” as described by Nagle and Morowitz (1978) and seen in the D pathway of Cytochrome *c* Oxidase (Lee et al., 2010, Henry et al., 2009) and the gramicidin A channel (Pomès and Roux, 1996); which act via a Grotthuss mechanism to transfer protons via hydrogen bonded water molecules (de Grotthuss, 2006, Henry et al., 2009). A potential scheme for this proton transfer is shown in Scheme 2, adapted from Pomès and Roux (1996)’s proton wire. The T1Cu is quickly reduced following exposure to X-rays (I) and free radicals are generated by hydrolysis of water (see Fig.3.20) (II). These radicals cause a series of proton transfers between water molecules, paired with electron transfer as demonstrated by laser flash photolysis on N90S mutations in AxNiR (Leferink et al., 2012a, Leferink et al., 2011, Leferink et al., 2012b), until it is released at the T2Cu site and binds the Asp bound nitrite (CuNOOH) complex (III). Donation of the second proton generates a water molecule and the product NO (IV).



Scheme 2

Based on the currently accepted reaction mechanism where nitrite binding leads to electron transfer between copper atoms (Strange et al., 1999, Hough et al., 2008b) you would expect the water chain to be better conserved with nitrite bound to the T2Cu than the native or product bound state; ready to transfer a proton to form water and NO. Datasets one and two show complete water chains with dual nitrite conformations present (Fig. 7.15a-b), however, W9 is not observed in datasets three or four (Fig. 7.15c-d) where nitrite has stabilised in the side-on conformation. Following this the presence of W9 is intermitant in the presence of dual conformation nitrite and NO ligands and not observed in the presence of the product alone. This is not entirely unexpected as a

proton has been donated by this stage so the water chain has carried out its intended function. Density for W5 in dataset seventeen (NO bound) has decreased significantly compared to dataset sixteen (nitrite/NO bound), as seen in Figure 7.15j-k. The average dose between these datasets saw an increase by 0.69 MGy from 11.04 MGy to 11.73 MG, not likely to cause any significant differences between datasets so far below the Henderson limit. This loss of density for W5 and complete loss of W9 may represent the relaxation of the water chain after proton donation. Following dataset seventeen there is a steady loss of density for W2, W4 and W5 (Fig. 7.15l-n), however, this becomes more difficult to judge as a true conformational change with increasing average dose.

Comparison of previously published *Ac*NiR structures in nitrite and NO bound states show a similarly conserved water chain between W1-7 as observed in serial data collections. Figure 7.16 shows superpositions of nitrite bound 2BWI (grey) and dataset one (purple) and NO bound 2BW5 (blue) against models from datasets seventeen (green) and eleven (orange). As expected, differences arise beyond W7. 2BWI shows an almost identical water chain with the exception of W8 which shows a third position not seen in dataset one (Fig. 7.16a). 2BW5 on the other hand displays several deviations from dataset seventeen with additional positons for W7, W8 and W9 (Fig. 7.16b). Of particular interest is the appearance of a fourth position for W8 (W8Pd), suggesting W8Pc is in fact a separate water molecule to W8Pa/b. The atomic resolution of this structure may explain why these waters were not observed in any other structures. The appearance of these extra waters suggests radiation damage may be too severe to allow study of the changes in the water chain with serial data collection. Although 2BWI's water chain differs in many ways to dataset seventeen, it does share many similarities with dataset eleven where W8Pc was first observed. However, 2BWI does still show additional positions for W7 and W9 not seen in dataset eleven (Fig. 7.16c). Comparison of five other nitrite reductases from other organisms shows varying levels of conservation of the water chain between copper atoms, (1) *Ac*NiR-NO_2 vs *Ax*NiR (1WA2/1OE1) with 7/6 waters in similar positions (Fig. 7.17a); (2) *Ac*NiR-NO_2 vs *Gt*NiR (3WNJ/4ZK8) with 7/7 conserved waters (Fig. 7.17b); (3) *Ac*NiR-NO_2 vs *Af*NiR (1SJM/1SNR/3H56) with 8/6/4 conserved waters (Fig. 7.17c); (4)) *Ac*NiR-NO_2

vs *Rs*NiR (1MZY/2DWS) with 6/2 conserved waters (Fig. 7.17d) and (5) *Ac*NiR-NO_2 vs *Rp*NiR (3ZIY) with 1 conserved water (Fig. 7.17e). The highest resolution structures available were used but some structures may lack water molecules as a result of lower resolution data. Overall comparison suggests a water chain is present in several nitrite reductases so may play a role on the reaction mechanism.

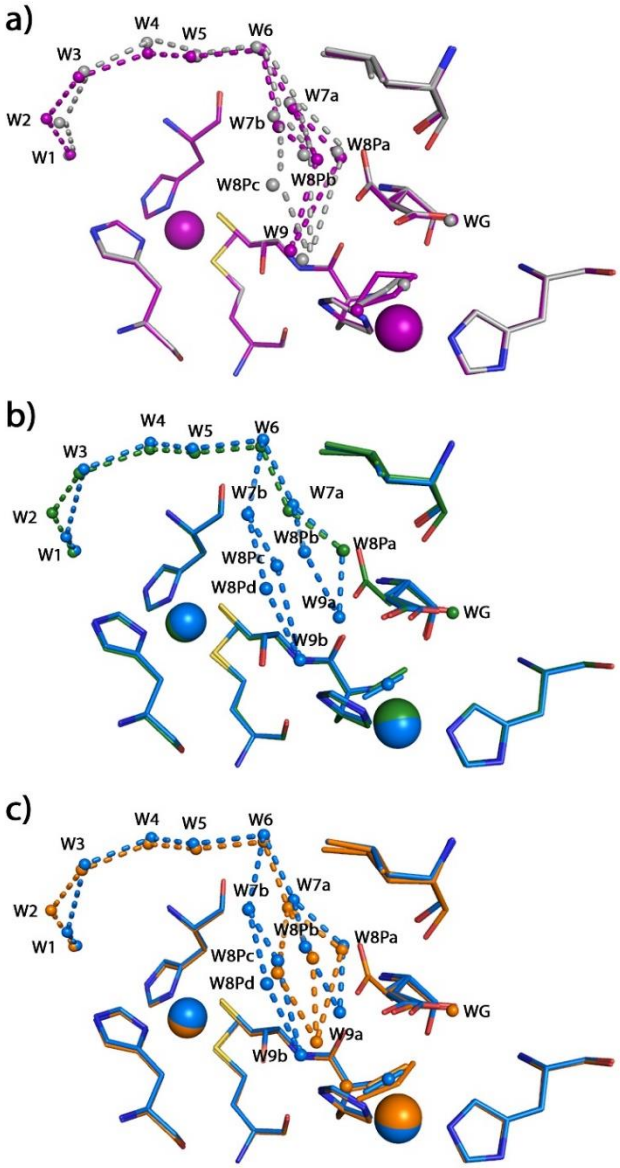


Fig. 7.16: Water chain superpositions of nitrite, nitrite/NO and NO bound AcNiR structures. a) Nitrite bound dataset one (purple) superposed with 2BWI (grey). b) NO bound dataset seventeen (green) superposed with 2BW5 (blue). c) Nitrite/NO bound dataset eleven (orange) superposed with NO bound 2BW4 (blue). Correspondingly coloured dashed lines represent potential electron transport pathways for structures as labelled above. All residues, nitrite and NO are displayed as sticks with residues coloured by element. Water molecules and copper atoms are displayed as spheres.

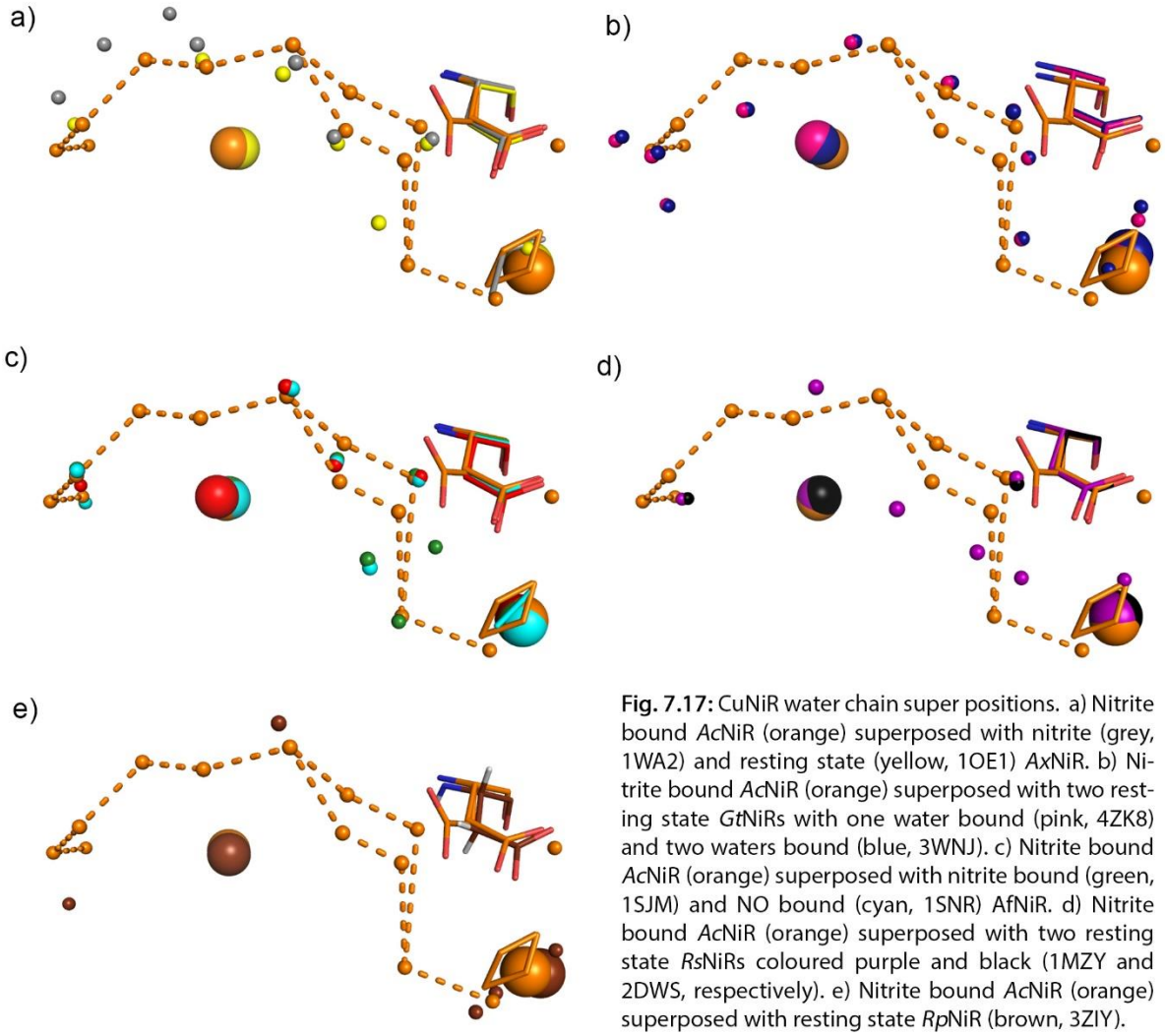


Fig. 7.17: CuNiR water chain super positions. a) Nitrite bound *Ac*NiR (orange) superposed with nitrite (grey, 1WA2) and resting state (yellow, 1OE1) *Ax*NiR. b) Nitrite bound *Ac*NiR (orange) superposed with two resting state *Gt*NiRs with one water bound (pink, 4ZK8) and two waters bound (blue, 3WNJ). c) Nitrite bound *Ac*NiR (orange) superposed with nitrite bound (green, 1SJM) and NO bound (cyan, 1SNR) AfNiR. d) Nitrite bound *Ac*NiR (orange) superposed with two resting state *Rs*NiRs coloured purple and black (1MZY and 2DWS, respectively). e) Nitrite bound *Ac*NiR (orange) superposed with resting state *Rp*NiR (brown, 3ZIY).

7.4. Conclusions

This study has used a single metallo-protein crystal at cryo temperatures diffracting to atomic resolution to probe the reaction mechanism of nitrite reductases by X-ray induced turnover of nitrite. The nature of this technique places certain requirements on the quality of the samples that can be analysed, such as near atomic resolution diffraction, resistance to radiation damage, high occupancy ligands and preferably a high symmetry space group for fast data collections. This serial data collection method has been made possible thanks to high throughput data collection developments at MX beamlines, particularly through the use of high speed detectors that allowed single datasets to be collected in 19 seconds. Automated data processing pipelines at Diamond Light

Source also allowed the “best” crystal to be easily identified for further analysis and made processing of the resulting 50 datasets a much less arduous task. However, use of automatic processing with minimal human intervention does have its downsides, as evident in Figure 7.9a where data appear to have been unnecessarily cut during processing. Overall the pros of automation easily outweigh the cons.

Increasing X-ray dose has been shown to result in the conversion of nitrite to NO at the T2Cu before incurring any significant radiation damage in the form of decreased resolution. Variation in the T2Cu-ligand density can clearly be seen over a number of datasets with two new conformations being observed; the dual conformation of nitrite representing the oscillation of the central N-atom before stabilising in the side on conformation and the appearance of the N02-NO-W-W model observed in datasets eleven and fourteen. Although modelling is not always straight forward this technique has allowed clear changes in the T2Cu site to be investigated. Furthermore, this study has shown evidence of the generation of an enzymatically stable side on symmetrical binding conformation for T2Cu-NO previously thought to be an end-on conformation by spectroscopic techniques. However, after the generation of NO, radiation damage becomes a more significant issue with maps becoming increasingly sparse and more difficult to interpret with confidence.

As well as evidence of substrate turnover in the active site and the side on conformation of NO this study has shown links between the conformation of Asp98 and the ligand present at the T2Cu active site. Although mixed occupation of the active site due to incomplete incorporation of ligands and incomplete turnover at any one time in the whole crystal make interpretation more difficult there appears to be a correlation between the presence of NO and an increase in the proximal conformation of Asp98. Study of the T1Cu-T2Cu water chain proved more difficult with radiation damage and loss of high resolution data limiting analysis of the full chain throughout the series, as shown by the additional waters seen in previously published atomic resolution structures.

However, this study has identified a highly conserved seven water chain which becomes more variable towards the T2Cu, likely as a result of the dual conformation of Asp98 which can impede the position of later waters in the chain. Similar water chains are observed in other nitrite reductases and mutation of residues in *Ax*NiR disrupting this water chain suggests this is the primary route for proton donation in CuNiRs. Based on this a mechanism for proton donation via the water chain has been proposed.

Future work to overcome the drawback of increased dose in later stages of the metallo-protein reactions could include the use of higher temperature (> 170K) or even room temperature data collections to allow faster X-ray induced substrate turnover. This would also remove the limited conformational flexibility incurred from use of cryogenic temperatures to give a more true to life picture if the reaction mechanism. *Ac*NiR crystals have been tested at room temperature and found to provide high resolution diffraction data without high levels of radiation damage. Use of ultra-fast detectors (> 100 Hz) and new developments in stable fast goniometry speeds up to 500°/s allowing a complete dataset to be collected in less than one second present an exciting opportunity for dose- and time-series kinetic crystallography studies of *Ac*NiR and other metallo-proteins.

7.5. Appendix

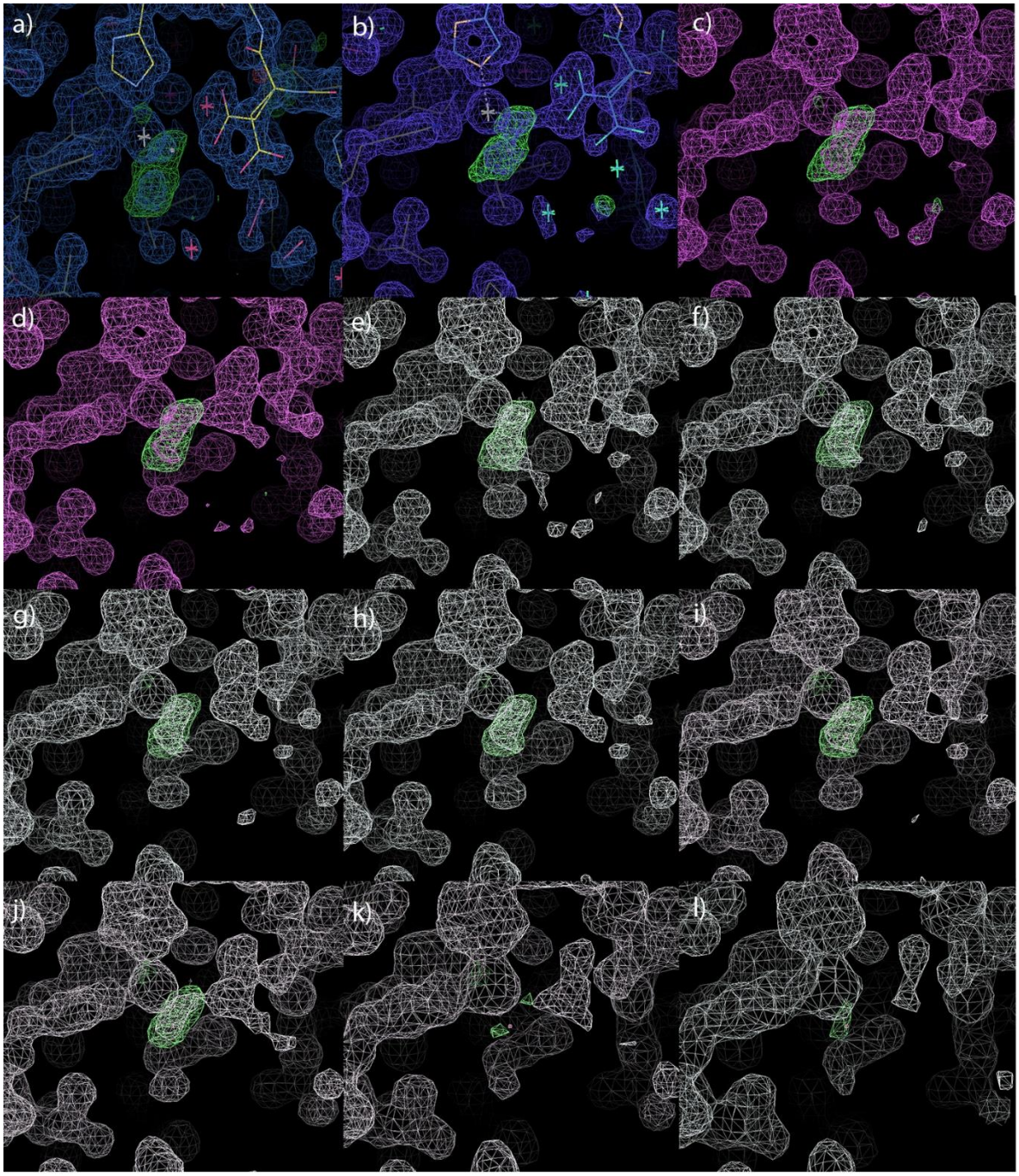


Fig. 7.A1: Nitrite soaked movie ligand difference density with $2F_o$-F_c density contoured to σ = 1 and F_o-F_c density contoured to σ = 3 and coloured in green. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3. (d) Dataset 4. (e) Dataset 11. (f) dataset 12. (g) Dataset 13. (h) Dataset 14. (i) Dataset 15. (j) Dataset 16. (k) Dataset 30. (l) Dataset 40.

Table 7.A1: Crystallization, crystallographic data collection and structure refinement statistics for nitrite soaked *Ac*NiR structures. Values in parentheses refer to the highest resolution bin.

Structure	**Dataset 1**	**Dataset 2**	**Dataset 3**	**Dataset 4**	**Dataset 11**
Data Collection					
Space Group	$P2_13$	$P2_13$	$P2_13$	$P2_13$	$P2_13$
Cell Axes (Å)	95.77	95.35	95.38	95.40	95.27
Angles (°)	90	90	90	90	90
Wavelength (Å)	0.92	0.92	0.92	0.92	0.92
Resolution (Å)	67.72-1.12	67.42-1.12	42.66-1.09	42.66-1.11	67.37-1.11
Unique Reflections (#)	116299 (8473)	122220 (8104)	116591 (8452)	110606 (8182)	110072 (8144)
Redundancy	4.0 (3.1)	4.0 (2.7)	4.0 (3.1)	4.0 (3.1)	4.0 (3.1)
R_{pim} (%)	2.4 (25.7)	2.4 (30.8)	2.4 (28.9)	2.5 (28.0)	2.2 (27.8)
I/σ(I)	18.0 (2.6)	17.2 (2.2)	17.1 (2.5)	16.9 (2.6)	19.1 (2.6)
Completeness (%)	92.82 (97.50)	96.9 (87.5)	97.6 (95.9)	97.28 (97.85)	97.17 (97.67)
CC(1/2)	0.999 (0.813)	0.999 (0.793)	0.999 (0.769)	0.999 (0.788)	0.999 (0.813)
Refinement					
Molecules/AU	1	1	1	1	1
R_{work}/R_{free} (%)	13.4/16.6	10.8/13.2	10.8/13.1	10.8/13.2	12.0/14.7
Rmsd Bond Length (Å)	0.029	0.035	0.033	0.034	0.034
Rmsd Bond Angles (°)	2.74	2.72	2.53	2.73	2.71
Average B factor (Å^2)	14.2	13.0	13.5	14.2	14.2
Ramachandran Statistics (%)					
Favoured Regions	97.34	97.48	97.49	97.39	97.43
Allowed Regions	2.66	2.52	2.51	2.61	2.66

Table 7.A2: Crystallization, crystallographic data collection and structure refinement statistics for nitrite soaked *Ac*NiR structures. Values in parentheses refer to the highest resolution bin.

Structure	Dataset 12	Dataset 13	Dataset 14	Dataset 15	Dataset 16
Data Collection					
Space Group	$P2_13$	$P2_13$	$P2_13$	$P2_13$	$P2_13$
Cell Axes (Å)	95.29	95.35	95.31	95.31	95.42
Angles (°)	90	90	90	90	90
Wavelength (Å)	0.92	0.92	0.92	0.92	0.92
Resolution (Å)	42.61-1.12	67.42-1.17	67.39-1.16	67.93-1.18	67.47-1.21
Unique Reflections (#)	107356 (7988)	102004 (7656)	96679 (7299)	91699 (6851)	85341 (6441)
Redundancy	4.1 (3.8)	4.1 (3.9)	4.1 (3.8)	4.1 (3.8)	4.1 (3.8)
R_{pim} (%)	2.3 (27.1)	2.3 (29.3)	2.2 (33.9)	2.2 (37.4)	2.2 (35.9)
I/σ(I)	19.4 (2.7)	18.8 (2.6)	19.0 (2.3)	18.8 (2.0)	19.1 (2.2)
Completeness (%)	97.7 (98.3)	97.7 (99.4)	97.6 (99.3)	97.5 (98.5)	97.5 (99.7)
CC(1/2)	0.999 (0.820)	0.999 (0.788)	0.999 (0.756)	0.999 (0.706)	0.999 (0.733)
Refinement					
Molecules/AU	1	1	1	1	1
R_{work}/R_{free} (%)	14.7/16.1	11.9/14.6	11.7/14.7	11.9/14.4	11.4/14.2
Rmsd Bond Length (Å)	0.031	0.033	0.031	0.034	0.031
Rmsd Bond Angles (°)	2.54	2.70	2.67	1.98	2.64
Average B factor (Å^2)	13.2	14.5	15.5	16.0	16.6
Ramachandran Statistics (%)					
Favoured Regions	97.67	97.34	97.01	98.01	98.01
Allowed Regions	2.33	2.66	2.99	1.99	1.99

Table 7.A3: Crystallization, crystallographic data collection and structure refinement statistics for nitrite soaked *Ac*NiR structures. Values in parentheses refer to the highest resolution bin.

Structure	Dataset 17	Dataset 30	Dataset 40
Data Collection			
Space Group	$P2_13$	$P2_13$	$P2_13$
Cell Axes (Å)	95.43	95.54	95.60
Angles (°)	90	90	90
Wavelength (Å)	0.92	0.92	0.92
Resolution (Å)	67.47-1.21	67.56-1.49	67.60-1.65
Unique Reflections (#)	85341 (6441)	45371 (3436)	31632 (2327)
Redundancy	4.1 (3.8)	4.2 (4.2)	4.2 (4.2)
R_{pim} (%)	2.2 (35.9)	2.3 (38.0)	2.3 (41.1)
I/σ(I)	19.1 (2.2)	18.0 (2.2)	17.8 (2.3)
Completeness (%)	97.5 (99.7)	96.1 (98.4)	94.5 (95.9)
CC(1/2)	(0.999) 0.727	0.999 (0.710)	0.999 (0.706)
Refinement			
Molecules/AU	1	1	1
R_{work}/R_{free} (%)	12.1/14.4	15.4/18.0	15.6 (18.1)
Rmsd Bond Length (Å)	0.033	0.031	0.023
Rmsd Bond Angles (°)	2.70	2.29	2.15
Average B factor (Å^2)	17.4	22.9	27.1
Ramachandran Statistics (%)			
Favoured Regions	98.01	98.05	98.07
Allowed Regions	1.99	1.95	1.93

Table 7.A4: T1Cu bond distances and angles for datasets 1-14 of the nitrite soaked *Ac*NiR structures.

Structure	1	2	3	4	11	12	13	14
Resolution (Å)	1.12	1.07	1.09	1.11	1.11	1.12	1.14	1.16
Bond length (Å)								
T1Cu-His95 Nδ1	2.05	2.02	2.03	2.02	2.03	1.96	2.02	2.01
T1Cu-His145 Nδ1	2.04	2.02	2.02	2.02	2.01	2.02	2.03	2.03
T1Cu-Cys136 Sγ	2.21	2.21	2.21	2.21	2.21	2.21	2.21	2.21
T1Cu-Met150 Sδ	2.52	2.49	2.5	2.5	2.5	2.49	2.5	2.49
Bond angle (°)								
His95 Nδ1-T1Cu- His145 Nδ1	100	101	100.9	101	100.9	101	101.2	101.2
His145 Nδ1- T1Cu-Cys136 Sγ	105.6	106.8	107.7	108.1	107.2	108	107.9	107.8
Cys136 Sγ- T1Cu-Met150 Sδ	107.2	107.3	107.2	107.2	107.3	107.1	107.5	107.5
Met150 Sδ- T1Cu-His95 Nδ1	87.3	85.9	85.9	86.4	86.6	86.5	86.6	87
Cys136 Sγ- T1Cu- His95 Nδ1	130.7	129.6	129.1	128.7	129.1	128.3	127.9	127.9
Met150 Sδ- T1Cu- His145 Nδ1	128.6	128	127.4	126.7	127.3	127.2	126.9	126.7

Table 7.A5: T1Cu bond distances and angles for datasets 15-40 of the nitrite soaked *Ac*NiR structures.

Structure	15	16	17	30	40
Resolution (Å)	1.18	1.21	1.23	1.49	1.65
Bond lengths (Å)					
T1Cu-His95 Nδ1	2.00	2.01	2.02	2.05	2.05
T1Cu-His145 Nδ1	2.03	2.04	2.05	2.06	2.04
T1Cu-Cys136 Sγ	2.21	2.22	2.22	2.22	2.24
T1Cu-Met150 Sδ	2.49	2.49	2.49	2.48	2.42
Bond angle (°)					
His95 Nδ1-T1Cu- His145 Nδ1	101.1	100.9	102.1	102.4	101.2
His145 Nδ1- T1Cu-Cys136 Sγ	108	108.4	107.9	108.1	109.3
Cys136 Sγ- T1Cu-Met150 Sδ	107.6	107.6	107.9	108.1	108.7
Met150 Sδ- T1Cu-His95 Nδ1	87.1	87	86.9	87.8	88.2
Cys136 Sγ- T1Cu- His95 Nδ1	127.9	128	127.7	127.4	127.8
Met150 Sδ- T1Cu- His145 Nδ1	126.4	125.9	125.3	123.9	121.9

Table 7.A6: T2Cu bond distances for datasets 1-13 of the nitrite soaked *Ac*NiR structures.

Structure	1	2	3	4	11	12	13
T2Cu-His100 Nε2	2.04	2.01	2	2.01	2.01	2	2.01
T2Cu-His135Nε2	2.08	2.06	2.04	2.05	2.06	2.05	2.05
T2Cu-His306 Nε2	2.02	2.01	2.04	2.03	2.03	2.01	2.04
T2Cu - His255Nε2	4.17	4.13	4.09	4.09	4.08	4.43	4.08
Cu – Nitrite A (O1/O2)	2.07/2.05	2.05/2.07	2.03/2.12	N/A	N/A	N/A	N/A
Cu – Nitrite B (O1/O2)	2.06/2.04	2.05/2.07	2.04/2.12	2.05/2.12	2.05/1.98	2.11/1.97	2.18/1.93
Cu-O/N (NO)	N/A	N/A	N/A	N/A	1.89/1.93	2.02/1.91	1.91/1.99
Cu-H_2O	N/A	N/A	N/A	N/A	2.49	2.5	2.85
Asp98 OD2-Nitrite O1 A (proximal)	2.38	2.3	2.37	2.1	N/A	N/A	N/A
Asp98 OD2-Nitrite O1 B (proximal)	2.41	2.32	2.21	N/A	2.22	2.19	1.85
Asp98-NO (proximal)	N/A	N/A	N/A	N/A	2.45	2.49	2.38
Asp98 occupancy A/B	0.5/0.5	0.6/0.4	0.6/0.4	0.6/0.4	0.5/0.5	0.5/0.5	0.5/0.5

Table 7.A7: T2Cu bond distances for datasets 14-40 of the nitrite soaked *Ac*NiR structures.

Structure	14	15	16	17	30	40
T2Cu-His100 Nε2	2.01	2.02	2.01	2	2.03	2.05
T2Cu-His135Nε2	2.04	2.03	2.04	2.04	2.04	2
T2Cu-His306 Nε2	2.05	2.05	2.06	2.06	2.04	2.05
T2Cu - His255Nε2	4.08	4.09	4.07	4.05	4.03	4.07
Cu – Nitrite A (O1/O2)	N/A	N/A	N/A	N/A	N/A	N/A
Cu – Nitrite B (O1/O2)	2.22/1.88	2.24/2.02	2.23/2.02	N/A	N/A	N/A
Cu-O/N (NO)	2.17/1.85	1.9/1.99	2.23/1.96	2.03/2.02	2.37/2.07	N/A
Cu-H_2O	2.51	2.35	2.54	N/A	N/A	2.09
Asp98 OD2-Nitrite O1 A (proximal)	N/A	N/A	N/A	N/A	N/A	N/A
Asp98 OD2-Nitrite O1 B (proximal)	1.81	1.7	1.77	N/A	N/A	N/A
Asp98-NO (proximal)	2.34	2.21	2.16	2.38	1.82	N/A
Asp98 occupancy A/B	0.5/0.5	0.5/0.5	0.65/0.35	0.65/0.35	0.6/0.0	0.6/0.0

Chapter 8 – Concluding Remarks

This thesis has investigated the viability of the use of a humidity controlled stream to preserve protein crystals to allow removal of excess surrounding solvent and found that additional sample handling and exposure to the humid air stream are not detrimental to the quality of diffraction data obtained from large populations of crystals analysed with robust statistical techniques. Applying the same sample transfer method and statistical analysis, large populations of crystals were produced with and without excess surrounding solvent and diffraction data were collected at standard MX wavelengths (~ 1 Å) and longer wavelengths (~2 Å). Data quality indicators demonstrated a protein dependent effect between long wavelength X-rays and excess surrounding solvent with *PA*3825-EAL showing no significant difference in data quality indicators while BPM2 showed a significant decrease in the quality of data collected at ~ 2 Å with excess surrounding solvent. This effect is likely to become more pronounced as longer wavelengths are achieved upon completion of I23 at Diamond.

The importance of a specialised long wavelength beamline has also been demonstrated through the unsuccessful attempts at solving the structure of *Ax*UDT by S-SAD phasing. In line with the studies on *Ax*UDT the structure of *Ax*ABH is presented in this thesis. The crystal structure of *Ax*ABH in several crystal forms has a well conserved α/β hydrolase fold with eight β-sheets flanked by five α-helices to position the catalytic triad in the active site which is formed by a novel fold of two helix turn helix motifs and one long α-helix. The specific hydrolase function conveyed by this novel fold remains unknown.

In addition to the long wavelength data quality experiments performed on *PA*3825-EAL this thesis investigated the regulation of EAL domain catalytic activity through structural and biochemical methods. The high resolution crystal structures of *PA*3825-EAL in its resting state, active state, substrate bound state and product bound state are presented here. Comparison of the active

monomeric and dimeric structures of *PA*3825-EAL with the inactive dimeric MucR-EAL points towards an EAL domain inhibition through the dimerisation interface at helix 8. The presence of bulky residues on helix 8 of *PA*3825-EAL prevent it forming the dimeric conformation observed in MucR-EAL and other EAL domain crystal structures where CdG bound in activating conditions. The pGpG bound structure of *PA*3825-EAL also shows the presence of a novel third metal binding site coordinated to the two oxygen atoms involved in formation of the CdG scissile phosphoester bond. The importance of this third metal in the active site is not yet clear but these structural data suggests it plays a role in activation of the transition state for bond cleavage.

Finally, this thesis demonstrates the use of serial data collection of atomic resolution data from a single nitrite reductase crystal soaked with nitrite at cryogenic temperatures. This work was made possible by significant advances on MX beamlines which allowed complete datasets to be collected in 19 seconds and the use of automated data processing software to assess the presence of nitrite at the T2Cu active site from terabytes of data. From this data a 45 dataset series from a single crystal diffracting to 1.07 Å at first were used to generate a movie of the reaction mechanism of *Ac*NiR as it reduced nitrite to NO. Over 45 datasets resolution decreased to 1.65 Å but the transition from nitrite to NO and the interaction of Asp98 with the bound ligand could be observed. A 30 Å water chain connecting the T2Cu active site to the surface was also identified with seven highly conserved waters which became more variable closer to the active site due to interactions with the dual conformation Asp98. This thesis presents a mechanism by which this proton wire donates the second proton in the reaction mechanism to form NO and water. This technique has the potential to investigate the reaction mechanism of a variety of metallo-proteins given initial high/atomic resolution diffraction data.

References

AARON, S. D., VANDEMHEEN, K. L., RAMOTAR, K., GIESBRECHT-LEWIS, T., TULLIS, E., FREITAG, A., PATERSON, N., JACKSON, M., LOUGHEED, M. D., DOWSON, C., KUMAR, V., FERRIS, W., CHAN, F., DOUCETTE, S. & FERGUSSON, D. 2010. Infection with transmissible strains of Pseudomonas aeruginosa and clinical outcomes in adults with cystic fibrosis. *JAMA,* 304**,** 2145-53.

ABEL, S., CHIEN, P., WASSMANN, P., SCHIRMER, T., KAEVER, V., LAUB, M. T., BAKER, T. A. & JENAL, U. 2011. Regulatory cohesion of cell cycle and cell differentiation through interlinked phosphorylation and second messenger networks. *Mol Cell,* 43**,** 550-60.

ABERGEL, C. 2004. Spectacular improvement of X-ray diffraction through fast desiccation of protein crystals. *Acta Crystallographica Section D,* 60**,** 1413-1416.

ABRAHAM, Z. H., LOWE, D. J. & SMITH, B. E. 1993. Purification and characterization of the dissimilatory nitrite reductase from Alcaligenes xylosoxidans subsp. xylosoxidans (N.C.I.M.B. 11015): evidence for the presence of both type 1 and type 2 copper centres. *Biochemical Journal,* 295**,** 587-593.

ABRAHAM, Z. H., SMITH, B. E., HOWES, B. D., LOWE, D. J. & EADY, R. R. 1997. pH-dependence for binding a single nitrite ion to each type-2 copper centre in the copper-containing nitrite reductase of Alcaligenes xylosoxidans. *Biochemical Journal,* 324**,** 511-516.

ADACHI, H., UMENA, Y., ENAMI, I., HENMI, T., KAMIYA, N. & SHEN, J. R. 2009. Towards structural elucidation of eukaryotic photosystem II: Purification, crystallization and preliminary X-ray diffraction analysis of photosystem II from a red alga. *Biochimica et Biophysica Acta - Bioenergetics,* 1787**,** 121-128.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L.-W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D,* 66**,** 213-221.

ADMAN, E. T., GODDEN, J. W. & TURLEY, S. 1995. The Structure of Copper-nitrite Reductase from Achromobacter cycloclastes at Five pH Values, with NO−2 Bound and with Type II Copper Depleted. *Journal of Biological Chemistry,* 270**,** 27458-27474.

AGARWAL, A., MACKENZIE, R. J., PIPPA, R., EIDE, C. A., ODDO, J., TYNER, J. W., SEARS, R., VITEK, M. P., ODERO, M. D., CHRISTENSEN, D. J. & DRUKER, B. J. 2014. Antagonism of SET using OP449 enhances the efficacy of tyrosine kinase inhibitors and overcomes drug resistance in myeloid leukemia. *Clin Cancer Res,* 20**,** 2092-103.

AHO, T. L., SANDHOLM, J., PELTOLA, K. J., MANKONEN, H. P., LILLY, M. & KOSKINEN, P. J. 2004. Pim-1 kinase promotes inactivation of the pro-apoptotic Bad protein by phosphorylating it on the Ser112 gatekeeper site. *FEBS Lett,* 571**,** 43-9.

ALESSI, D. R., CAUDWELL, F. B., ANDJELKOVIC, M., HEMMINGS, B. A. & COHEN, P. 1996. Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p70 S6 kinase. *FEBS Letters,* 399**,** 333-338.

ALEXANDER, P. & LETT, J. T. 1967. *Comprehensive Biochemistry,* 27**,** 267.

ALKIRE, R. W., DUKE, N. E. C. & ROTELLA, F. J. 2008. Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion. *Journal of Applied Crystallography,* 41**,** 1122-1133.

ALKIRE, R. W., ROTELLA, F. J. & DUKE, N. E. C. 2013. Testing commercial protein crystallography sample mounting loops for movement in a cold-stream. *Journal of Applied Crystallography,* 46**,** 525-536.

ALLAN, E. G., KANDER, M. C., CARMICHAEL, I. & GARMAN, E. F. 2013. To scavenge or not to scavenge, that is STILL the question. *J Synchrotron Radiat,* 20**,** 23-36.

ALONSO, M. D., LOMAKO, J., LOMAKO, W. M. & WHELAN, W. J. 1995. A new look at the biogenesis of glycogen. *FASEB J,* 9**,** 1126-37.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

AMIKAM, D. & GALPERIN, M. Y. 2006. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics,* 22**,** 3-6.

AN, S., WU, J. & ZHANG, L. H. 2010. Modulation of Pseudomonas aeruginosa biofilm dispersal by a cyclic-Di-GMP phosphodiesterase with a putative hypoxia-sensing domain. *Appl Environ Microbiol,* 76**,** 8160-73.

ANDRABI, S., GJOERUP, O. V., KEAN, J. A., ROBERTS, T. M. & SCHAFFHAUSEN, B. 2007. Protein phosphatase 2A regulates life and death decisions via Akt in a context-dependent manner. *Proc Natl Acad Sci U S A,* 104**,** 19011-6.

ANTONYUK, S. V. & HOUGH, M. A. 2011. Monitoring and validating active site redox states in protein crystals. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* 1814**,** 778-784.

ANTONYUK, S. V., STRANGE, R. W., SAWERS, G., EADY, R. R. & HASNAIN, S. S. 2005. Atomic resolution structures of resting-state, substrate- and product-complexed Cu-nitrite reductase provide insight into catalytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 12041-12046.

ARMSTRONG, D. S., NIXON, G. M., CARZINO, R., BIGHAM, A., CARLIN, J. B., ROBINS-BROWNE, R. M. & GRIMWOOD, K. 2002. Detection of a widespread clone of Pseudomonas aeruginosa in a pediatric cystic fibrosis clinic. *Am J Respir Crit Care Med,* 166**,** 983-7.

ARNDT, U. 1984. Optimum X-ray wavelength for protein crystallography. *Journal of Applied Crystallography,* 17**,** 118-119.

ARNDT, U. W., CROWTHER, R. A. & MALLETT, J. F. 1968. A computer-linked cathode-ray tube microdensitometer for x-ray crystallography. *J Sci Instrum,* 1**,** 510-6.

AVERILL, B. A. 1996. Dissimilatory Nitrite and Nitric Oxide Reductases. *Chem Rev,* 96**,** 2951-2964.

AWAD, W., SVENSSON BIRKEDAL, G., THUNNISSEN, M. M., MANI, K. & LOGAN, D. T. 2013. Improvements in the order, isotropy and electron density of glypican-1 crystals by controlled dehydration. *Acta Crystallogr D Biol Crystallogr,* 69**,** 2524-33.

BARENDS, T. R., HARTMANN, E., GRIESE, J. J., BEITLICH, T., KIRIENKO, N. V., RYJENKOV, D. A., REINSTEIN, J., SHOEMAN, R. L., GOMELSKY, M. & SCHLICHTING, I. 2009. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature,* 459**,** 1015-8.

BARKEN, K. B., PAMP, S. J., YANG, L., GJERMANSEN, M., BERTRAND, J. J., KLAUSEN, M., GIVSKOV, M., WHITCHURCH, C. B., ENGEL, J. N. & TOLKER-NIELSEN, T. 2008. Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in Pseudomonas aeruginosa biofilms. *Environ Microbiol,* 10**,** 2331-43.

BEITLICH, T., KUHNEL, K., SCHULZE-BRIESE, C., SHOEMAN, R. L. & SCHLICHTING, I. 2007. Cryoradiolytic reduction of crystalline heme proteins: analysis by UV-Vis spectroscopy and X-ray crystallography. *J Synchrotron Radiat,* 14**,** 11-23.

BELLINI, D., CALY, D. L., MCCARTHY, Y., BUMANN, M., AN, S. Q., DOW, J. M., RYAN, R. P. & WALSH, M. A. 2014. Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. *Mol Microbiol,* 91**,** 26-38.

BENKERT, P., BIASINI, M. & SCHWEDE, T. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics,* 27**,** 343-50.

BIOU, V., BOSECKE, P., BOIS, J.-M., BRANDOLIN, G., KAHN, R., MAS, C., NAUTON, L., NURY, H., PEBAY-PEYROULA, E., VICAT, J. & STUHRMANN, H. 2005. X-ray spectroscopy and X-ray diffraction at wavelengths near the K-absorption edge of phosphorus. *Journal of Synchrotron Radiation,* 12**,** 402-409.

BOCKELMAN, C., LASSUS, H., HEMMES, A., LEMINEN, A., WESTERMARCK, J., HAGLUND, C., BUTZOW, R. & RISTIMAKI, A. 2011. Prognostic role of CIP2A expression in serous ovarian cancer. *Br J Cancer,* 105**,** 989-95.

BONOFIGLIO, L., GARCIA, E. & MOLLERACH, M. 2005. Biochemical characterization of the pneumococcal glucose 1-phosphate uridylyltransferase (GalU) essential for capsule biosynthesis. *Curr Microbiol,* 51**,** 217-21.

BORLEE, B. R., GOLDMAN, A. D., MURAKAMI, K., SAMUDRALA, R., WOZNIAK, D. J. & PARSEK, M. R. 2010. Pseudomonas aeruginosa uses a cyclic-di-GMP-regulated adhesin to reinforce the biofilm extracellular matrix. *Mol Microbiol,* 75**,** 827-42.

BOULANGER, M. J., KUKIMOTO, M., NISHIYAMA, M., HORINOUCHI, S. & MURPHY, M. E. P. 2000. Catalytic Roles for Two Water Bridged Residues (Asp-98 and His-255) in the Active Site of Copper-containing Nitrite Reductase. *Journal of Biological Chemistry,* 275**,** 23957-23964.

BOULANGER, M. J. & MURPHY, M. E. P. 2003. Directing the mode of nitrite binding to a copper-containing nitrite reductase from Alcaligenes faecalis S-6: Characterization of an active site isoleucine. *Protein Science : A Publication of the Protein Society,* 12**,** 248-256.

BOUTET, S., LOMB, L., WILLIAMS, G. J., BARENDS, T. R. M., AQUILA, A., DOAK, R. B., WEIERSTALL, U., DEPONTE, D. P., STEINBRENER, J., SHOEMAN, R. L., MESSERSCHMIDT, M., BARTY, A., WHITE, T. A., KASSEMEYER, S., KIRIAN, R. A., SEIBERT, M. M., MONTANEZ, P. A., KENNEY, C., HERBST, R., HART, P., PINES, J., HALLER, G., GRUNER, S. M., PHILIPP, H. T., TATE, M. W., HROMALIK, M., KOERNER, L. J., VAN BAKEL, N., MORSE, J., GHONSALVES, W., ARNLUND, D., BOGAN, M. J., CALEMAN, C., FROMME, R., HAMPTON, C. Y., HUNTER, M. S., JOHANSSON, L. C., KATONA, G., KUPITZ, C., LIANG, M., MARTIN, A. V., NASS, K., REDECKE, L., STELLATO, F., TIMNEANU, N., WANG, D., ZATSEPIN, N. A., SCHAFER, D., DEFEVER, J., NEUTZE, R., FROMME, P., SPENCE, J. C. H., CHAPMAN, H. N. & SCHLICHTING, I. 2012. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science,* 337**,** 362-364.

BOUWMAN, A. F., BEUSEN, A. H. W., GRIFFIOEN, J., VAN GROENIGEN, J. W., HEFTING, M. M., OENEMA, O., VAN PUIJENBROEK, P. J. T. M., SEITZINGER, S., SLOMP, C. P. & STEHFEST, E. 2013. Global trends and uncertainties in terrestrial denitrification and N(2)O emissions. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20130112.

BOWLER, M. W., MONTGOMERY, M. G., LESLIE, A. G. & WALKER, J. E. 2006a. How azide inhibits ATP hydrolysis by the F-ATPases. *Proc Natl Acad Sci U S A,* 103**,** 8646-9.

BOWLER, M. W., MONTGOMERY, M. G., LESLIE, A. G. & WALKER, J. E. 2006b. Reproducible improvements in order and diffraction limit of crystals of bovine mitochondrial F(1)-ATPase by controlled dehydration. *Acta Crystallogr D Biol Crystallogr,* 62**,** 991-5.

BOWLER, M. W., MONTGOMERY, M. G., LESLIE, A. G. & WALKER, J. E. 2007. Ground state structure of F1-ATPase from bovine heart mitochondria at 1.9 A resolution. *J Biol Chem,* 282**,** 14238-42.

BOWLER, M. W., MONTGOMERY, M. G., LESLIE, A. G. W. & WALKER, J. E. 2006c. Reproducible improvements in order and diffraction limit of crystals of bovine mitochondrial F1-ATPase by controlled dehydration. *Acta Crystallographica Section D,* 62**,** 991-995.

BOX, H. C. & BUDZINSKI, E. E. 1971. Oxidation and Reduction of Amino Acids by Ionizing Radiation. *The Journal of Chemical Physics,* 55**,** 2446-2449.

BOYD, C. D. & O'TOOLE, G. A. 2012. Second messenger regulation of biofilm formation: breakthroughs in understanding c-di-GMP effector systems. *Annu Rev Cell Dev Biol,* 28**,** 439-62.

BRAGG, W. H. & BRAGG, W. L. 1913. *The Reflection of X-rays by Crystals*.

BRENNER, S., HEYES, D. J., HAY, S., HOUGH, M. A., EADY, R. R., HASNAIN, S. S. & SCRUTTON, N. S. 2009. Demonstration of Proton-coupled Electron Transfer in the Copper-containing Nitrite Reductases. *The Journal of Biological Chemistry,* 284**,** 25973-25983.

BROCKHAUSER, S., DI MICHIEL, M., MCGEEHAN, J. E., MCCARTHY, A. A. & RAVELLI, R. B. G. 2008. X-ray tomographic reconstruction of macromolecular samples. *Journal of Applied Crystallography,* 41**,** 1057-1066.

BROENNIMANN, C., EIKENBERRY, E. F., HENRICH, B., HORISBERGER, R., HUELSEN, G., POHL, E., SCHMITT, B., SCHULZE-BRIESE, C., SUZUKI, M., TOMIZAKI, T., TOYOKAWA, H. & WAGNER, A. 2006a. The PILATUS 1M detector. *J Synchrotron Radiat,* 13**,** 120-30.

BROENNIMANN, C., EIKENBERRY, E. F., HENRICH, B., HORISBERGER, R., HUELSEN, G., POHL, E., SCHMITT, B., SCHULZE-BRIESE, C., SUZUKI, M., TOMIZAKI, T., TOYOKAWA, H. & WAGNER, A. 2006b. The PILATUS 1M detector. *Journal of Synchrotron Radiation,* 13**,** 120-130.

BUCHAN, D. W., WARD, S. M., LOBLEY, A. E., NUGENT, T. C., BRYSON, K. & JONES, D. T. 2010. Protein annotation and modelling servers at University College London. *Nucleic Acids Res,* 38**,** W563-8.

BURKHARD, P., STETEFELD, J. & STRELKOV, S. V. 2001. Coiled coils: a highly versatile protein folding motif. *Trends in Cell Biology,* 11**,** 82-88.

BURMEISTER, W. 2000. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica Section D,* 56**,** 328-341.

BURNS, J. L., GIBSON, R. L., MCNAMARA, S., YIM, D., EMERSON, J., ROSENFELD, M., HIATT, P., MCCOY, K., CASTILE, R., SMITH, A. L. & RAMSEY, B. W. 2001. Longitudinal Assessment of Pseudomonas aeruginosa in Young Children with Cystic Fibrosis. *Journal of Infectious Diseases,* 183**,** 444-452.

CALY, D. L., BELLINI, D., WALSH, M. A., DOW, J. M. & RYAN, R. P. 2014. Targeting Cyclic di-GMP Signalling: a Strategy to Control Biofilm Formation? *Curr Pharm Des*.

CDC. 2014. *Pseudomonas aeruginosa in Healthcare Settings* [Online]. http://www.cdc.gov/hai/organisms/pseudomonas.html: Centre for Disease Control and Prevention [Accessed 30th April 2015].

CEGIELSKA, A., SHAFFER, S., DERUA, R., GORIS, J. & VIRSHUP, D. M. 1994. Different oligomeric forms of protein phosphatase 2A activate and inhibit simian virus 40 DNA replication. *Mol Cell Biol,* 14**,** 4616-23.

CHAMBLESS, J. D., HUNT, S. M. & STEWART, P. S. 2006. A Three-Dimensional Computer Model of Four Hypothetical Mechanisms Protecting Biofilms from Antimicrobials. *Applied and Environmental Microbiology,* 72**,** 2005-2013.

CHAN, C., PAUL, R., SAMORAY, D., AMIOT, N. C., GIESE, B., JENAL, U. & SCHIRMER, T. 2004. Structural basis of activity and allosteric control of diguanylate cyclase. *Proc Natl Acad Sci U S A,* 101**,** 17084-9.

CHANA, M., TRIPET, B. P., MANT, C. T. & HODGES, R. S. 2002. The Role of Unstructured Highly Charged Regions on the Stability and Specificity of Dimerization of Two-Stranded α-Helical Coiled-Coils: Analysis of the Neck-Hinge Region of the Kinesin-like Motor Protein Kif3A. *Journal of Structural Biology,* 137**,** 206-219.

CHANG, C. W., COUNAGO, R. L., WILLIAMS, S. J., BODEN, M. & KOBE, B. 2012. Crystal structure of rice importin-alpha and structural basis of its interaction with plant-specific nuclear localization signals. *Plant Cell,* 24**,** 5074-88.

CHANG, C. W., WILLIAMS, S. J., COUNAGO, R. M. & KOBE, B. 2014. Structural basis of interaction of bipartite nuclear localization signal from Agrobacterium VirD2 with rice importin-alpha. *Mol Plant,* 7**,** 1061-4.

CHEN, M. W., KOTAKA, M., VONRHEIN, C., BRICOGNE, G., RAO, F., CHUAH, M. L., SVERGUN, D., SCHNEIDER, G., LIANG, Z. X. & LESCAR, J. 2012. Structural insights into the regulatory mechanism of the response regulator RocR from Pseudomonas aeruginosa in cyclic Di-GMP signaling. *J Bacteriol,* 194**,** 4837-46.

CHEN, V. B., ARENDALL, W. B., 3RD, HEADD, J. J., KEEDY, D. A., IMMORMINO, R. M., KAPRAL, G. J., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2010. MolProbity: all-atom

structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr,* 66**,** 12-21.
CHENG, K., SMYTH, R. L., GOVAN, J. R., DOHERTY, C., WINSTANLEY, C., DENNING, N., HEAF, D. P., VAN SAENE, H. & HART, C. A. 1996. Spread of beta-lactam-resistant Pseudomonas aeruginosa in a cystic fibrosis clinic. *Lancet,* 348**,** 639-42.
CHIANG, P. & BURROWS, L. L. 2003. Biofilm Formation by Hyperpiliated Mutants of Pseudomonas aeruginosa. *Journal of Bacteriology,* 185**,** 2374-2378.
CHO, U. S. & XU, W. 2007. Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature,* 445**,** 53-7.
CHOTIYARNWONG, P., STEWART-JONES, G. B., TARRY, M. J., DEJNIRATTISAI, W., SIEBOLD, C., KOCH, M., STUART, D. I., HARLOS, K., MALASIT, P., SCREATON, G., MONGKOLSAPAYA, J. & JONES, E. Y. 2007. Humidity control as a strategy for lattice optimization applied to crystals of HLA-A*1101 complexed with variant peptides from dengue virus. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 63**,** 386-92.
CHRISTEN, M., CHRISTEN, B., FOLCHER, M., SCHAUERTE, A. & JENAL, U. 2005. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J Biol Chem,* 280**,** 30829-37.
CHRISTEN, M., KULASEKARA, H. D., CHRISTEN, B., KULASEKARA, B. R., HOFFMAN, L. R. & MILLER, S. I. 2010. Asymmetrical Distribution of the Second Messenger c-di-GMP upon Bacterial Cell Division. *Science,* 328**,** 1295-1297.
CICCONE, M., CALIN, G. A. & PERROTTI, D. 2015. From the Biology of PP2A to the PADs for Therapy of Hematologic Malignancies. *Front Oncol,* 5**,** 21.
CIPRIANI, F., OUML, WER, M., LANDRET, C., ZANDER, U., FELISAZ, F., AACUTE, RQUEZ, J., EACUTE, ANTONIO, CIPRIANI, F., OUML, WER, M., LANDRET, C., ZANDER, U., FELISAZ, F., AACUTE, RQUEZ, J., EACUTE & ANTONIO 2012. CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallographica Section D: Biological Crystallography,* 68**,** 1393.
COATES, J. C. 2003. Armadillo repeat proteins: beyond the animal kingdom. *Trends in Cell Biology,* 13**,** 463-471.
COHEN, P. 1990. The structure and regulation of protein phosphatases. *Adv Second Messenger Phosphoprotein Res,* 24**,** 230-5.
COLE, C., BARBER, J. D. & BARTON, G. J. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res,* 36**,** W197-201.
COME, C., LAINE, A., CHANRION, M., EDGREN, H., MATTILA, E., LIU, X., JONKERS, J., IVASKA, J., ISOLA, J., DARBON, J. M., KALLIONIEMI, O., THEZENAS, S. & WESTERMARCK, J. 2009. CIP2A is associated with human breast cancer aggressivity. *Clin Cancer Res,* 15**,** 5092-100.
CONTI, E. & KURIYAN, J. 2000. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin α. *Structure,* 8**,** 329-338.
CONTI, E., UY, M., LEIGHTON, L., BLOBEL, G. & KURIYAN, J. 1998. Crystallographic Analysis of the Recognition of a Nuclear Localization Signal by the Nuclear Import Factor Karyopherin α. *Cell,* 94**,** 193-204.
COSTERTON, J. W., STEWART, P. S. & GREENBERG, E. P. 1999. Bacterial Biofilms: A Common Cause of Persistent Infections. *Science,* 284**,** 1318-1322.
COWAN, J. A. & NAVE, C. 2008. The optimum conditions to collect X-ray data from very small samples. *Journal of Synchrotron Radiation,* 15**,** 458-462.
COWTAN, K. 2012. Completion of autobuilt protein models using a database of protein fragments. *Acta Crystallogr D Biol Crystallogr,* 68**,** 328-35.
CRAMER, P., BUSHNELL, D. A., FU, J., GNATT, A. L., MAIER-DAVIS, B., THOMPSON, N. E., BURGESS, R. R., EDWARDS, A. M., DAVID, P. R. & KORNBERG, R. D. 2000. Architecture of RNA Polymerase II and Implications for the Transcription Mechanism. *Science,* 288**,** 640-649.

DANIELS, D. L., EKLOF SPINK, K. & WEIS, W. I. 2001. β-catenin: molecular plasticity and drug design. *Trends in Biochemical Sciences,* 26**,** 672-678.
DAUTER, Z. & ADAMIAK, D. A. 2001. Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy. *Acta Crystallographica Section D,* 57**,** 990-995.
DAUTER, Z., DAUTER, M., DE LA FORTELLE, E., BRICOGNE, G. & SHELDRICK, G. M. 1999. Can anomalous signal of sulfur become a tool for solving protein crystal structures?1. *Journal of Molecular Biology,* 289**,** 83-92.
DAVIDSON, A. G., CHILVERS, M. A. & LILLQUIST, Y. P. 2012. Effects of a Pseudomonas aeruginosa eradication policy in a cystic fibrosis clinic. *Curr Opin Pulm Med,* 18**,** 615-21.
DAVIES, C. E., HILL, K. E., WILSON, M. J., STEPHENS, P., HILL, C. M., HARDING, K. G. & THOMAS, D. W. 2004. Use of 16S Ribosomal DNA PCR and Denaturing Gradient Gel Electrophoresis for Analysis of the Microfloras of Healing and Nonhealing Chronic Venous Leg Ulcers. *Journal of Clinical Microbiology,* 42**,** 3549-3557.
DAVIES, D. G. & MARQUES, C. N. H. 2009. A Fatty Acid Messenger Is Responsible for Inducing Dispersion in Microbial Biofilms. *Journal of Bacteriology,* 191**,** 1393-1403.
DE GROTTHUSS, C. J. T. 2006. Memoir on the decomposition of water and of the bodies that it holds in solution by means of galvanic electricity. *Biochimica et Biophysica Acta - Bioenergetics,* 1757**,** 871-875.
DENG, X., DAVIDSON, W. S. & THOMPSON, T. B. 2012. Improving the diffraction of apoA-IV crystals through extreme dehydration. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 68**,** 105-10.
DIEDERICHS, K. & KARPLUS, P. A. 1997. Improved R-factors for diffraction data analysis in macromolecular crystallography.
DIEGELMANN, R. F. 2003. Excessive neutrophils characterize chronic pressure ulcers. *Wound Repair Regen,* 11**,** 490-5.
DJINOVIC CARUGO, K., HELLIWELL, J. R., STUHRMANN, H. & WEISS, M. S. 2005. Softer and soft X-rays in macromolecular crystallography. *Journal of Synchrotron Radiation,* 12**,** 410-419.
DODD, F. E., HASNAIN, S. S., ABRAHAM, Z. H. L., EADY, R. R. & SMITH, B. E. 1997. Structures of a Blue-Copper Nitrite Reductase and its Substrate-Bound Complex. *Acta Crystallographica Section D,* 53**,** 406-418.
DODD, F. E., VAN BEEUMEN, J., EADY, R. R. & HASNAIN, S. S. 1998. X-ray structure of a blue-copper nitrite reductase in two crystal forms. The nature of the copper sites, mode of substrate binding and recognition by redox partner1. *Journal of Molecular Biology,* 282**,** 369-382.
DONG, Q. Z., WANG, Y., DONG, X. J., LI, Z. X., TANG, Z. P., CUI, Q. Z. & WANG, E. H. 2011. CIP2A is overexpressed in non-small cell lung cancer and correlates with poor prognosis. *Ann Surg Oncol,* 18**,** 857-65.
DOSZTANYI, Z., CSIZMOK, V., TOMPA, P. & SIMON, I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics,* 21**,** 3433-4.
DOUTCH, J., HOUGH, M. A., HASNAIN, S. S. & STRANGE, R. W. 2012. Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. *Journal of Synchrotron Radiation,* 19**,** 19-29.
DOWD, S., SUN, Y., SECOR, P., RHOADS, D., WOLCOTT, B., JAMES, G. & WOLCOTT, R. 2008. Survey of bacterial diversity in chronic wounds using Pyrosequencing, DGGE, and full ribosome shotgun sequencing. *BMC Microbiology,* 8**,** 43.
DUEHOLM, M. S., SONDERGAARD, M. T., NILSSON, M., CHRISTIANSEN, G., STENSBALLE, A., OVERGAARD, M. T., GIVSKOV, M., TOLKER-NIELSEN, T., OTZEN, D. E. & NIELSEN, P. H. 2013. Expression of Fap amyloids in Pseudomonas aeruginosa, P. fluorescens, and P. putida results in aggregation and increased biofilm formation. *Microbiologyopen,* 2**,** 365-82.
EIBEN, K. & TAUB, I. A. 1967. Solvated Electron Spectrum in Irradiated Ice. *Nature,* 216**,** 782-783.
EICHHORN, P. J., CREYGHTON, M. P. & BERNARDS, R. 2009. Protein phosphatase 2A regulatory subunits and cancer. *Biochim Biophys Acta,* 1795**,** 1-15.

EIFERT, C. & POWERS, R. S. 2012. From cancer genomes to oncogenic drivers, tumour dependencies and therapeutic targets. *Nat Rev Cancer,* 12**,** 572-8.

EISENBERG, D., LUTHY, R. & BOWIE, J. U. 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol,* 277**,** 396-404.

EL GHACHI, M., BOUHSS, A., BARRETEAU, H., TOUZE, T., AUGER, G., BLANOT, D. & MENGIN-LECREULX, D. 2006. Colicin M exerts its bacteriolytic effect via enzymatic degradation of undecaprenyl phosphate-linked peptidoglycan precursors. *J Biol Chem,* 281**,** 22761-72.

EL OMARI, K., IOURIN, O., KADLEC, J., FEARN, R., HALL, D. R., HARLOS, K., GRIMES, J. M. & STUART, D. I. 2014. Pushing the limits of sulfur SAD phasing: de novo structure solution of the N-terminal domain of the ectodomain of HCV E1. *Acta Crystallographica Section D,* 70**,** 2197-2203.

ELLIS, M. J., BUFFEY, S. G., HOUGH, M. A. & HASNAIN, S. S. 2008. On-line optical and X-ray spectroscopies with crystallography: an integrated approach for determining metalloprotein structures in functionally well defined states. *J Synchrotron Radiat,* 15**,** 433-9.

ELLIS, M. J., PRUDÊNCIO, M., DODD, F. E., STRANGE, R. W., SAWERS, G., EADY, R. R. & HASNAIN, S. S. 2002. Biochemical and crystallographic studies of the Met144Ala, Asp92Asn and His254Phe mutants of the nitrite reductase from Alcaligenes xylosoxidans provide insight into the enzyme mechanism1. *Journal of Molecular Biology,* 316**,** 51-64.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

ENGEL, M., HOFFMANN, T., WAGNER, L., WERMANN, M., HEISER, U., KIEFERSAUER, R., HUBER, R., BODE, W., DEMUTH, H. U. & BRANDSTETTER, H. 2003. The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. *Proc Natl Acad Sci U S A,* 100**,** 5063-8.

ESNOUF, R. M., REN, J., GARMAN, E. F., SOMERS, D. O. N., ROSS, C. K., JONES, E. Y., STAMMERS, D. K. & STUART, D. I. 1998. Continuous and Discontinuous Changes in the Unit Cell of HIV-1 Reverse Transcriptase Crystals on Dehydration. *Acta Crystallographica Section D,* 54**,** 938-953.

EVANS, G., ALIANELLI, L., BURT, M., WAGNER, A. & SAWHNEY, K. J. S. 2007. Diamond Beamline 124: A Flexible Instrument for Macromolecular Micro-crystallography. *AIP Conference Proceedings,* 879**,** 836-839.

EVANS, P. R. & MURSHUDOV, G. N. 2013. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr,* 69**,** 1204-14.

FARAGGI, M., FERRADINI, C., HOUEE-LEVIN, C. & KLAPPER, M. H. 1995. *Radiation Research 1895-1995,* 2**,** 241-248.

FAZLI, M., BJARNSHOLT, T., KIRKETERP-MOLLER, K., JORGENSEN, A., ANDERSEN, C. B., GIVSKOV, M. & TOLKER-NIELSEN, T. 2011. Quantitative analysis of the cellular inflammatory response against biofilm bacteria in chronic wounds. *Wound Repair Regen,* 19**,** 387-91.

FRANCOIS, P., SCHRENZEL, J., STOERMAN-CHOPARD, C., FAVRE†, H., HERRMANN, M., FOSTER, T. J., LEW, D. P. & VAUDAUX, P. 2000. Identification of plasma proteins adsorbed on hemodialysis tubing that promote Staphylococcus aureus adhesion. *Journal of Laboratory and Clinical Medicine,* 135**,** 32-42.

FREDERIKSEN, B., KOCH, C. & HOIBY, N. 1999. Changing epidemiology of Pseudomonas aeruginosa infection in Danish cystic fibrosis patients (1974-1995). *Pediatr Pulmonol,* 28**,** 159-66.

FRIEDMAN, L. & KOLTER, R. 2004. Genes involved in matrix formation in Pseudomonas aeruginosa PA14 biofilms. *Mol Microbiol,* 51**,** 675-90.

GABY, W. L. & HADLEY, C. 1957. PRACTICAL LABORATORY TEST FOR THE IDENTIFICATION OF PSEUDOMONAS AERUGINOSA. *Journal of Bacteriology,* 74**,** 356-358.

GALPERIN, M. Y. 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol,* 5**,** 35.

GALPERIN, M. Y., GAIDENKO, T. A., MULKIDJANIAN, A. Y., NAKANO, M. & PRICE, C. W. 2001. MHYT, a new integral membrane sensor domain. *FEMS Microbiol Lett,* 205**,** 17-23.

GARMAN, E. 2003a. 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Curr Opin Struct Biol,* 13**,** 545-51.

GARMAN, E. 2003b. 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Current Opinion in Structural Biology,* 13**,** 545-551.

GARMAN, E. 2010. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D,* 66**,** 339-351.

GARMAN, E. F. & OWEN, R. L. 2006. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallographica Section D,* 62**,** 32-47.

GARMAN, E. F. & WEIK, M. 2015. Radiation damage to macromolecules: kill or cure? *Journal of Synchrotron Radiation,* 22**,** 195-200.

GATI, C., BOURENKOV, G., KLINGE, M., REHDERS, D., STELLATO, F., OBERTHUR, D., YEFANOV, O., SOMMER, B. P., MOGK, S., DUSZENKO, M., BETZEL, C., SCHNEIDER, T. R., CHAPMAN, H. N. & REDECKE, L. 2014. Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ,* 1**,** 87-94.

GENEVAUX, P., BAUDA, P., DUBOW, M. S. & OUDEGA, B. 1999. Identification of Tn10 insertions in the rfaG, rfaP, and galU genes involved in lipopolysaccharide core biosynthesis that affect Escherichia coli adhesion. *Arch Microbiol,* 172**,** 1-8.

GERMILLER, J. A., EL-KASHLAN, H. K. & SHAH, U. K. 2005. Chronic Pseudomonas infections of cochlear implants. *Otol Neurotol,* 26**,** 196-201.

GHOSH, S., DEY, A., USOV, O. M., SUN, Y., GRIGORYANTS, V. M., SCHOLES, C. P. & SOLOMON, E. I. 2007. Resolution of the Spectroscopy versus Crystallography Issue for NO Intermediates of Nitrite Reductase from Rhodobacter sphaeroides. *Journal of the American Chemical Society,* 129**,** 10310-10311.

GJERMANSEN, M., RAGAS, P., STERNBERG, C., MOLIN, S. & TOLKER-NIELSEN, T. 2005. Characterization of starvation-induced dispersion in Pseudomonas putida biofilms. *Environ Microbiol,* 7**,** 894-906.

GJODSBOL, K., CHRISTENSEN, J. J., KARLSMARK, T., JORGENSEN, B., KLEIN, B. M. & KROGFELT, K. A. 2006. Multiple bacterial species reside in chronic wounds: a longitudinal study. *Int Wound J,* 3**,** 225-31.

GODDEN, J., TURLEY, S., TELLER, D., ADMAN, E., LIU, M., PAYNE, W. & LEGALL, J. 1991. The 2.3 angstrom X-ray structure of nitrite reductase from Achromobacter cycloclastes. *Science,* 253**,** 438-442.

GOLDENBERG, O., EREZ, E., NIMROD, G. & BEN-TAL, N. 2009. The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res,* 37**,** D323-7.

GOVAN, J. R. & DERETIC, V. 1996. Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. *Microbiological Reviews,* 60**,** 539-574.

GRAY, J. S., BIRMINGHAM, J. M. & FENTON, J. I. 2010. Got black swimming dots in your cell culture? Identification of Achromobacter as a novel cell culture contaminant. *Biologicals : journal of the International Association of Biological Standardization,* 38**,** 273-277.

GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. 1954. *The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method*.

GRINTER, R., JOSTS, I., ZETH, K., ROSZAK, A. W., MCCAUGHEY, L. C., COGDELL, R. J., MILNER, J. J., KELLY, S. M., BYRON, O. & WALKER, D. 2014. Structure of the atypical bacteriocin pectocin M2 implies a novel mechanism of protein uptake. *Mol Microbiol,* 93**,** 234-46.

GRINTER, R., ROSZAK, A. W., COGDELL, R. J., MILNER, J. J. & WALKER, D. 2012. The crystal structure of the lipid II-degrading bacteriocin syringacin M suggests unexpected evolutionary relationships between colicin M-like bacteriocins. *J Biol Chem,* 287**,** 38876-88.

GRUBBS, F. E. 1950. Sample Criteria for Testing Outlying Observations. 27-58.

GUGGENBICHLER, J. P., ASSADIAN, O., BOESWALD, M. & KRAMER, A. 2011. Incidence and clinical implication of nosocomial infections associated with implantable biomaterials - catheters, ventilator-associated pneumonia, urinary tract infections. *GMS Krankenhhyg Interdiszip,* 6**,** Doc18.

GUPTA, V., GUPTA, R. K., KHARE, G., SALUNKE, D. M., SUROLIA, A. & TYAGI, A. K. 2010. Structural Ordering of Disordered Ligand-Binding Loops of Biotin Protein Ligase into Active Conformations as a Consequence of Dehydration. *PLoS ONE,* 5**,** e9222.

HALL-STOODLEY, L., HU, F., GIESEKE, A. & ET AL. 2006. DIrect detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *JAMA,* 296**,** 202-211.

HALL-STOODLEY, L. & STOODLEY, P. 2005. Biofilm formation and dispersal and the transmission of human pathogens. *Trends in Microbiology,* 13**,** 7-10.

HAMMER, B. K. & BASSLER, B. L. 2009. Distinct Sensory Pathways in Vibrio cholerae El Tor and Classical Biotypes Modulate Cyclic Dimeric GMP Levels To Control Biofilm Formation. *Journal of Bacteriology,* 191**,** 169-177.

HAN, S., PARK, K., BAE, B.-N., KIM, K., KIM, H.-J., KIM, Y.-D. & KIM, H.-Y. 2003. E2F1 Expression is Related with the Poor Survival of Lymph Node-positive Breast Cancer Patients Treated with Fluorouracil, Doxorubicin and Cyclophosphamide. *Breast Cancer Research and Treatment,* 82**,** 11-16.

HANAHAN, D. & WEINBERG, R. A. 2000. The hallmarks of cancer. *Cell,* 100**,** 57-70.

HANAHAN, D. & WEINBERG, ROBERT A. 2011. Hallmarks of Cancer: The Next Generation. *Cell,* 144**,** 646-674.

HARKNESS, R. E. & BRAUN, V. 1989. Colicin M inhibits peptidoglycan biosynthesis by interfering with lipid carrier recycling. *J Biol Chem,* 264**,** 6177-82.

HAY, I. D., REMMINGHORST, U. & REHM, B. H. 2009. MucR, a novel membrane-associated regulator of alginate biosynthesis in Pseudomonas aeruginosa. *Appl Environ Microbiol,* 75**,** 1110-20.

HENDERSON, R. 1990. *Cryo-Protection of Protein Crystals against Radiation Damage in Electron and X-Ray Diffraction*.

HENDRICKSON, W. A. & TEETER, M. M. 1981. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature,* 290**,** 107-113.

HENGGE, R. 2009. Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol,* 7**,** 263-73.

HENRICH, S., CAMERON, A., BOURENKOV, G. P., KIEFERSAUER, R., HUBER, R., LINDBERG, I., BODE, W. & THAN, M. E. 2003. The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat Struct Biol,* 10**,** 520-6.

HENRY, R. M., YU, C. H., RODINGER, T. & POMES, R. 2009. Functional hydration and conformational gating of proton uptake in cytochrome c oxidase. *J Mol Biol,* 387**,** 1165-85.

HENTZER, M., EBERL, L. & GIVSKOV, M. 2005. Transcriptome analysis of Pseudomonas aeruginosa biofilm development: anaerobic respiration and iron limitation. *Biofilms,* 2**,** 37-61.

HERAS, B., EDELING, M. A., BYRIEL, K. A., JONES, A., RAINA, S. & MARTIN, J. L. 2003. Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure,* 11**,** 139-145.

HERAS, B. & MARTIN, J. L. 2005. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D,* 61**,** 1173-1180.

HILDEBRAND, A., REMMERT, M., BIEGERT, A. & SODING, J. 2009. Fast and accurate automatic structure prediction with HHpred. *Proteins,* 77 Suppl 9**,** 128-32.

HIROKI, A., PIMBLOTT, S. M. & LAVERNE, J. A. 2002. Hydrogen Peroxide Production in the Radiolysis of Water with High Radical Scavenger Concentrations. *The Journal of Physical Chemistry A,* 106**,** 9352-9358.

HOIBY, N. & PEDERSEN, S. S. 1989. Estimated risk of cross-infection with Pseudomonas aeruginosa in Danish cystic fibrosis patients. *Acta Paediatr Scand,* 78**,** 395-404.

HOLMQUIST, M. 2000. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr Protein Pept Sci,* 1**,** 209-35.

HOLTON, J. M. & FRANKEL, K. A. 2010. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallogr D Biol Crystallogr,* 66**,** 393-408.
HOPE, H., FROLOW, F., VON BOHLEN, K., MAKOWSKI, I., KRATKY, C., HALFON, Y., DANZ, H., WEBSTER, P., BARTELS, K. S., WITTMANN, H. G. & ET AL. 1989. Cryocrystallography of ribosomal particles. *Acta Crystallogr B,* 45 (Pt 2)**,** 190-9.
HOUGH, M. A., ANTONYUK, S. V., STRANGE, R. W., EADY, R. R. & HASNAIN, S. S. 2008a. Crystallography with online optical and X-ray absorption spectroscopies demonstrates an ordered mechanism in copper nitrite reductase. *J Mol Biol,* 378**,** 353-61.
HOUGH, M. A., ANTONYUK, S. V., STRANGE, R. W., EADY, R. R. & HASNAIN, S. S. 2008b. Crystallography with Online Optical and X-ray Absorption Spectroscopies Demonstrates an Ordered Mechanism in Copper Nitrite Reductase. *Journal of Molecular Biology,* 378**,** 353-361.
HOUGH, M. A., ELLIS, M. J., ANTONYUK, S., STRANGE, R. W., SAWERS, G., EADY, R. R. & SAMAR HASNAIN, S. 2005. High Resolution Structural Studies of Mutants Provide Insights into Catalysis and Electron Transfer Processes in Copper Nitrite Reductase. *Journal of Molecular Biology,* 350**,** 300-309.
HU, N. J., IWATA, S., CAMERON, A. D. & DREW, D. 2011. Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. *Nature,* 478**,** 408-11.
HUBER, A. H., NELSON, W. J. & WEIS, W. I. 1997. Three-Dimensional Structure of the Armadillo Repeat Region of β-Catenin. *Cell,* 90**,** 871-882.
INGWERSEN, J., AKTAS, O., KUERY, P., KIESEIER, B., BOYKO, A. & HARTUNG, H. P. 2012. Fingolimod in multiple sclerosis: mechanisms of action and clinical efficacy. *Clin Immunol,* 142**,** 15-24.
INOUE, T., GOTOWDA, M., DELIGEER, KATAOKA, K., YAMAGUCHI, K., SUZUKI, S., WATANABE, H., GOHOW, M. & KAI, Y. 1998. Type 1 Cu Structure of Blue Nitrite Reductase from Alcaligenes xylosoxidans GIFU 1051 at 2.05 Å Resolution: Comparison of Blue and Green Nitrite Reductases. *Journal of Biochemistry,* 124**,** 876-879.
IRVING, H. & WILLIAMS, R. J. P. 1953. 637. The stability of transition-metal complexes. *Journal of the Chemical Society (Resumed)***,** 3192-3210.
ISHIYAMA, N., LEE, S.-H., LIU, S., LI, G.-Y., SMITH, M. J., REICHARDT, L. F. & IKURA, M. 2010. Dynamic and Static Interactions between p120 Catenin and E-Cadherin Regulate the Stability of Cell-Cell Adhesion. *Cell,* 141**,** 117-128.
JANSSENS, V., LONGIN, S. & GORIS, J. 2008. PP2A holoenzyme assembly: in cauda venenum (the sting is in the tail). *Trends Biochem Sci,* 33**,** 113-21.
JETTEN, M. S. M. 2008. The microbial nitrogen cycle. *Environmental Microbiology,* 10**,** 2903-2909.
JONES, P., BINNS, D., CHANG, H.-Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S.-Y., LOPEZ, R. & HUNTER, S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics,* 30**,** 1236-1240.
JUHAS, M., EBERL, L. & TUMMLER, B. 2005. Quorum sensing: the power of cooperation in the world of Pseudomonas. *Environ Microbiol,* 7**,** 459-71.
JUNTTILA, M. R., PUUSTINEN, P., NIEMELA, M., AHOLA, R., ARNOLD, H., BOTTZAUW, T., ALA-AHO, R., NIELSEN, C., IVASKA, J., TAYA, Y., LU, S. L., LIN, S., CHAN, E. K., WANG, X. J., GRENMAN, R., KAST, J., KALLUNKI, T., SEARS, R., KAHARI, V. M. & WESTERMARCK, J. 2007. CIP2A inhibits PP2A in human malignancies. *Cell,* 130**,** 51-62.
JUNTTILA, M. R. & WESTERMARCK, J. 2008. Mechanisms of MYC stabilization in human malignancies. *Cell Cycle,* 7**,** 592-6.
KABSCH, W. 2010a. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr,* 66**,** 133-44.
KABSCH, W. 2010b. XDS. *Acta Crystallogr D Biol Crystallogr,* 66**,** 125-32.

KALININ, Y., KMETKO, J., BARTNIK, A., STEWART, A., GILLILAN, R., LOBKOVSKY, E. & THORNE, R. 2005. A new sample mounting technique for room-temperature macromolecular crystallography. *Journal of Applied Crystallography,* 38**,** 333-339.

KAPPOS, L., RADUE, E. W., O'CONNOR, P., POLMAN, C., HOHLFELD, R., CALABRESI, P., SELMAJ, K., AGOROPOULOU, C., LEYK, M., ZHANG-AUBERSON, L. & BURTIN, P. 2010. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med,* 362**,** 387-401.

KARPLUS, P. A. & DIEDERICHS, K. 2012. Linking crystallographic model and data quality. *Science,* 336**,** 1030-3.

KAZMIERCZAK, B. I., LEBRON, M. B. & MURRAY, T. S. 2006. Analysis of FimX, a phosphodiesterase that governs twitching motility in Pseudomonas aeruginosa. *Mol Microbiol,* 60**,** 1026-43.

KEEGAN, R. M. & WINN, M. D. 2007. Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr D Biol Crystallogr,* 63**,** 447-57.

KEEGAN, R. M. & WINN, M. D. 2008. MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr D Biol Crystallogr,* 64**,** 119-24.

KEROSUO, L., FOX, H., PERALA, N., AHLQVIST, K., SUOMALAINEN, A., WESTERMARCK, J., SARIOLA, H. & WARTIOVAARA, K. 2010. CIP2A increases self-renewal and is linked to Myc in neural progenitor cells. *Differentiation,* 80**,** 68-77.

KHANNA, A., BOCKELMAN, C., HEMMES, A., JUNTTILA, M. R., WIKSTEN, J. P., LUNDIN, M., JUNNILA, S., MURPHY, D. J., EVAN, G. I., HAGLUND, C., WESTERMARCK, J. & RISTIMAKI, A. 2009. MYC-dependent regulation and prognostic role of CIP2A in gastric cancer. *J Natl Cancer Inst,* 101**,** 793-805.

KHANNA, A., PIMANDA, J. E. & WESTERMARCK, J. 2013. Cancerous inhibitor of protein phosphatase 2A, an emerging human oncoprotein and a potential cancer therapy target. *Cancer Res,* 73**,** 6548-53.

KIEFERSAUER, R., STETEFELD, J., GOMIS-RUTH, F. X., ROMAO, M. J., LOTTSPEICH, F. & HUBER, R. 1996. Protein-Crystal Density by Volume Measurement and Amino-Acid Analysis. *Journal of Applied Crystallography,* 29**,** 311-317.

KIEFERSAUER, R., THAN, M. E., DOBBEK, H., GREMER, L., MELERO, M., STROBL, S., DIAS, J. M., SOULIMANE, T. & HUBER, R. 2000. A novel free-mounting system for protein crystals: transformation and improvement of diffraction power by accurately controlled humidity changes. *Journal of Applied Crystallography,* 33**,** 1223-1230.

KIMURA, A., MOUNTZOUROS, K. T., RELMAN, D. A., FALKOW, S. & COWELL, J. L. 1990. Bordetella pertussis filamentous hemagglutinin: evaluation as a protective antigen and colonization factor in a mouse respiratory infection model. *Infect Immun,* 58**,** 7-16.

KITAGO, Y., WATANABE, N. & TANAKA, I. 2005a. Structure determination of a novel protein by sulfur SAD using chromium radiation in combination with a new crystal-mounting method. *Acta Crystallogr D Biol Crystallogr,* 61**,** 1013-21.

KITAGO, Y., WATANABE, N. & TANAKA, I. 2010a. Semi-automated protein crystal mounting device for the sulfur single-wavelength anomalous diffraction method. *Journal of Applied Crystallography,* 43**,** 341-346.

KITAGO, Y., WATANABE, N., TANAKA, I., KITAGO, Y., WATANABE, N. & TANAKA, I. 2010b. Semi‐automated protein crystal mounting device for the sulfur single‐wavelength anomalous diffraction method. *Journal of Applied Crystallography,* 43**,** 341.

KITAGO, Y., WATANABE, N., TANAKA, I., KITAGO, Y., WATANABE, N., TANAKA, I., KITAGO, Y., WATANABE, N., TANAKA, I., KITAGO, Y., WATANABE, N., TANAKA, I., KITAGO, Y., WATANABE, N. & TANAKA, I. 2005b. Structure determination of a novel protein by sulfur SAD using chromium radiation in combination with a new crystal-mounting method. *Acta Crystallographica Section D: Biological Crystallography,* 61**,** 1013.

KITANO, H., ADACHI, H., MURAKAMI, A., MATSUMURA, H., TAKANO, K., INOUE, T., MORI, Y., OWA, S. & SASAKI, T. 2004. Protein Crystal Processing Using a Deep-UV Laser. *Japanese Journal of Applied Physics,* 43**,** L73.

KITANO, H., MATSUMURA, H., ADACHI, H., MURAKAMI, S., TAKANO, S., INOUE, T., MORI, Y., DOI, M. & SASAKI, T. 2005a. Protein Cryocrystallography Using Laser-Processed Crystal. *Japanese Journal of Applied Physics,* 44**,** L54.

KITANO, H., MURAKAMI, S., ADACHI, H., MATSUMURA, H., TAKANO, K., INOUE, T., MORI, Y., DOI, M. & SASAKI, T. 2005b. Processing of membrane protein crystal using ultraviolet laser irradiation. *Journal of Bioscience and Bioengineering,* 100**,** 50-53.

KITATANI, T., ADACHI, H., SUGIYAMA, S., MATSUMURA, H., MURAI, R., TAKAHASHI, Y., MURAKAMI, S., INOUE, T., MORI, Y. & TAKANO, K. 2008a. Evaluation and Improvement of a Technique to Manipulate Protein Crystals in Solution. *Japanese Journal of Applied Physics,* 47**,** 8995-8997.

KITATANI, T., ADACHI, H., SUGIYAMA, S., MATSUMURA, H., MURAI, R., TAKAHASHI, Y., MURAKAMI, S., INOUE, T., MORI, Y. & TAKANO, K. 2009. A Manipulating Tool for Protein Microcrystals in Solution Using Adhesive Materials. *Japanese Journal of Applied Physics,* 48**,** 118001.

KITATANI, T., SUGIYAMA, S., MATSUMURA, H., ADACHI, H., YOSHIKAWA, H. Y., MAKI, S., MURAKAMI, S., INOUE, T., MORI, Y. & TAKANO, T. 2008b. New Technique of Manipulating a Protein Crystal Using Adhesive Material. *Applied Physics Express,* 1**,** 037002.

KLAUSEN, M., AAES-JORGENSEN, A., MOLIN, S. & TOLKER-NIELSEN, T. 2003. Involvement of bacterial migration in the development of complex multicellular structures in Pseudomonas aeruginosa biofilms. *Mol Microbiol,* 50**,** 61-8.

KRISSINEL, E. & HENRICK, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D,* 60**,** 2256-2268.

KRISSINEL, E. & HENRICK, K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol,* 372**,** 774-97.

KUKIMOTO, M., NISHIYAMA, M., MURPHY, M. E., TURLEY, S., ADMAN, E. T., HORINOUCHI, S. & BEPPU, T. 1994. X-ray structure and site-directed mutagenesis of a nitrite reductase from Alcaligenes faecalis S-6: roles of two copper atoms in nitrite reduction. *Biochemistry,* 33**,** 5246-52.

KUO, A., BOWLER, M. W., ZIMMER, J., ANTCLIFF, J. F. & DOYLE, D. A. 2003. Increasing the diffraction limit and internal order of a membrane protein crystal by dehydration. *Journal of Structural Biology,* 141**,** 97-102.

KYRIELEIS, O. J. P., GOETTIG, P., KIEFERSAUER, R., HUBER, R. & BRANDSTETTER, H. 2005. Crystal structures of the tricorn interacting factor F3 from Thermoplasma acidophilum, a zinc aminopeptidase in three different conformations. *Journal of Molecular Biology,* 349**,** 787-800.

LACROIX, L. B., SHADLE, S. E., WANG, Y., AVERILL, B. A., HEDMAN, B., HODGSON, K. O. & SOLOMON, E. I. 1996. Electronic Structure of the Perturbed Blue Copper Site in Nitrite Reductase: Spectroscopic Properties, Bonding, and Implications for the Entatic/Rack State. *Journal of the American Chemical Society,* 118**,** 7755-7768.

LAINE, A., SIHTO, H., COME, C., ROSENFELDT, M. T., ZWOLINSKA, A., NIEMELA, M., KHANNA, A., CHAN, E. K., KAHARI, V. M., KELLOKUMPU-LEHTINEN, P. L., SANSOM, O. J., EVAN, G. I., JUNTTILA, M. R., RYAN, K. M., MARINE, J. C., JOENSUU, H. & WESTERMARCK, J. 2013. Senescence sensitivity of breast cancer cells is defined by positive feedback loop between CIP2A and E2F1. *Cancer Discov,* 3**,** 182-97.

LANGE, M., WEILAND, B. & HUTTERMANN, J. 1995. Influence of electron scavengers on the radical formation in thymidine-5'-monophosphate and DNA in frozen aqueous solution and glasses. *Int J Radiat Biol,* 68**,** 475-86.

LEE, H. J., SVAHN, E., SWANSON, J. M. J., LEPP, H., VOTH, G. A., BRZEZINSKI, P. & GENNIS, R. B. 2010. The Intricate Role of Water in Proton Transport through Cytochrome c Oxidase. *Journal of the American Chemical Society,* 132**,** 16225-16239.

LEE, J., PARK, C. K., PARK, J. O., LIM, T., PARK, Y. S., LIM, H. Y., LEE, I., SOHN, T. S., NOH, J. H., HEO, J. S., KIM, S., LIM, D. H., KIM, K.-M. & KANG, W. K. 2008. Impact of E2F-1 Expression on Clinical Outcome of Gastric Adenocarcinoma Patients with Adjuvant Chemoradiation Therapy. *Clinical Cancer Research,* 14**,** 82-88.

LEFERINK, N. G. H., EADY, R. R., HASNAIN, S. S. & SCRUTTON, N. S. 2012a. Laser-flash photolysis indicates that internal electron transfer is triggered by proton uptake by Alcaligenes xylosoxidans copper-dependent nitrite reductase. *FEBS Journal,* 279**,** 2174-2181.

LEFERINK, N. G. H., HAN, C., ANTONYUK, S. V., HEYES, D. J., RIGBY, S. E. J., HOUGH, M. A., EADY, R. R., SCRUTTON, N. S. & HASNAIN, S. S. 2011. Proton-Coupled Electron Transfer in the Catalytic Cycle of Alcaligenes xylosoxidans Copper-Dependent Nitrite Reductase. *Biochemistry,* 50**,** 4121-4131.

LEFERINK, N. G. H., PUDNEY, C. R., BRENNER, S., HEYES, D. J., EADY, R. R., SAMAR HASNAIN, S., HAY, S., RIGBY, S. E. J. & SCRUTTON, N. S. 2012b. Gating mechanisms for biological electron transfer: Integrating structure with biophysics reveals the nature of redox control in cytochrome P450 reductase and copper-dependent nitrite reductase. *FEBS Letters,* 586**,** 578-584.

LIBBY, E. & AVERILL, B. A. 1992. Evidence that the Type 2 copper centers are the site of nitrite reduction by Achromobacter cycloclastes nitrite reductase. *Biochemical and Biophysical Research Communications,* 187**,** 1529-1535.

LIU, Q., DAHMANE, T., ZHANG, Z., ASSUR, Z., BRASCH, J., SHAPIRO, L., MANCIA, F. & HENDRICKSON, W. A. 2012. Structures from Anomalous Diffraction of Native Biological Macromolecules. *Science,* 336**,** 1033-1037.

LIU, Z. J., VYSOTSKI, E. S., CHEN, C. J., ROSE, J. P., LEE, J. & WANG, B. C. 2000. Structure of the Ca2+-regulated photoprotein obelin at 1.7 A resolution determined directly from its sulfur substructure. *Protein Sci,* 9**,** 2085-93.

LOPEZ-JARAMILLO, F. J., GARCIA-RUIZ, J. M., GAVIRA, J. A. & OTALORA, F. 2001. Crystallization and cryocrystallography inside X-ray capillaries. *Journal of Applied Crystallography,* 34**,** 365-370.

LOVELL, S. C., DAVIS, I. W., ARENDALL, W. B., 3RD, DE BAKKER, P. I., WORD, J. M., PRISANT, M. G., RICHARDSON, J. S. & RICHARDSON, D. C. 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins,* 50**,** 437-50.

LOVERING, A. L., CAPENESS, M. J., LAMBERT, C., HOBLEY, L. & SOCKETT, R. E. 2011. The structure of an unconventional HD-GYP protein from Bdellovibrio reveals the roles of conserved residues in this class of cyclic-di-GMP phosphodiesterases. *MBio,* 2.

MA, L., JACKSON, K. D., LANDRY, R. M., PARSEK, M. R. & WOZNIAK, D. J. 2006. Analysis of Pseudomonas aeruginosa Conditional Psl Variants Reveals Roles for the Psl Polysaccharide in Adhesion and Maintaining Biofilm Structure Postattachment. *Journal of Bacteriology,* 188**,** 8213-8221.

MADSEN, S. M., WESTH, H., DANIELSEN, L. & ROSDAHL, V. T. 1996. Bacterial colonization and healing of venous leg ulcers. *APMIS,* 104**,** 895-9.

MAHENTHIRALINGAM, E. & SPEERT, D. P. 1995. Nonopsonic phagocytosis of Pseudomonas aeruginosa by macrophages and polymorphonuclear leukocytes requires the presence of the bacterial flagellum. *Infection and Immunity,* 63**,** 4519-4523.

MALKIN, R. & MALMSTRÖM, B. G. 2006. The State and Function of Copper in Biological Systems. *Advances in Enzymology and Related Areas of Molecular Biology.* John Wiley & Sons, Inc.

MANN, H. B. & WHITNEY, D. R. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Statist* 18**,** 11.

MARCHOT, P. & CHATONNET, A. 2012. Enzymatic activity and protein interactions in alpha/beta hydrolase fold proteins: moonlighting versus promiscuity. *Protein Pept Lett,* 19**,** 132-43.

MARTÍNEZ-PASTOR, J. C., MUÑOZ-MAHAMUD, E., VILCHEZ, F., GARCÍA-RAMIRO, S., BORI, G., SIERRA, J., MARTÍNEZ, J. A., FONT, L., MENSA, J. & SORIANO, A. 2009. Outcome of Acute Prosthetic

Joint Infections Due to Gram-Negative Bacilli Treated with Open Debridement and Retention of the Prosthesis. *Antimicrobial Agents and Chemotherapy,* 53**,** 4772-4777.
MASON, J. M. & ARNDT, K. M. 2004. Coiled Coil Domains: Stability, Specificity, and Biological Implications. *ChemBioChem,* 5**,** 170-176.
MATHIASEN, D. P., EGEBJERG, C., ANDERSEN, S. H., RAFN, B., PUUSTINEN, P., KHANNA, A., DAUGAARD, M., VALO, E., TUOMELA, S., BOTTZAUW, T., NIELSEN, C. F., WILLUMSEN, B. M., HAUTANIEMI, S., LAHESMAA, R., WESTERMARCK, J., JAATTELA, M. & KALLUNKI, T. 2012. Identification of a c-Jun N-terminal kinase-2-dependent signal amplification cascade that regulates c-Myc levels in ras transformation. *Oncogene,* 31**,** 390-401.
MATTHEWS, B. 1976. X-ray crystallographic studies of proteins. *Annual Review of Physical Chemistry,* 27**,** 493-493.
MATTHEWS, B. W. 1968. Solvent content of protein crystals. *J Mol Biol,* 33**,** 491-7.
MAZZORANA, M., SANCHEZ-WEATHERBY, J., SANDY, J., LOBLEY, C. M. C. & SORENSEN, T. 2014. An evaluation of adhesive sample holders for advanced crystallographic experiments. *Acta Crystallographica Section D,* 70**,** 2390-2400.
MCCOURT, P., GALLO-EBERT, C., GONGHONG, Y., JIANG, Y. & NICKELS, J. T., JR. 2013. PP2A(Cdc55) regulates G1 cyclin stability. *Cell Cycle,* 12**,** 1201-10.
MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40**,** 658-674.
MCGUFFIN, L. J., BRYSON, K. & JONES, D. T. 2000. The PSIPRED protein structure prediction server. *Bioinformatics,* 16**,** 404-5.
MEENTS, A., WAGNER, A., SCHNEIDER, R., PRADERVAND, C., POHL, E. & SCHULZE-BRIESE, C. 2007. Reduction of X-ray-induced radiation damage of macromolecular crystals by data collection at 15 K: a systematic study. *Acta Crystallogr D Biol Crystallogr,* 63**,** 302-9.
MILLWARD, T. A., ZOLNIEROWICZ, S. & HEMMINGS, B. A. 1999. Regulation of protein kinase cascades by protein phosphatase 2A. *Trends Biochem Sci,* 24**,** 186-91.
MINASOV, G., PADAVATTAN, S., SHUVALOVA, L., BRUNZELLE, J. S., MILLER, D. J., BASLE, A., MASSA, C., COLLART, F. R., SCHIRMER, T. & ANDERSON, W. F. 2009. Crystal structures of YkuI and its complex with second messenger cyclic Di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains. *J Biol Chem,* 284**,** 13174-84.
MIZIANTY, M. J. & KURGAN, L. 2011. Sequence-based prediction of protein crystallization, purification and production propensity. *Bioinformatics,* 27**,** i24-33.
MOLLERACH, M. & GARCIA, E. 2000. The galU gene of Streptococcus pneumoniae that codes for a UDP-glucose pyrophosphorylase is highly polymorphic and suitable for molecular typing and phylogenetic studies. *Gene,* 260**,** 77-86.
MOLLERACH, M., LOPEZ, R. & GARCIA, E. 1998. Characterization of the galU gene of Streptococcus pneumoniae encoding a uridine diphosphoglucose pyrophosphorylase: a gene essential for capsular polysaccharide biosynthesis. *J Exp Med,* 188**,** 2047-56.
MORGAN, R., KOHN, S., HWANG, S. H., HASSETT, D. J. & SAUER, K. 2006. BdlA, a chemotaxis regulator essential for biofilm dispersion in Pseudomonas aeruginosa. *J Bacteriol,* 188**,** 7335-43.
MORISHITA, E. C., MURAYAMA, K., KATO-MURAYAMA, M., ISHIZUKA-KATSURA, Y., TOMABECHI, Y., HAYASHI, T., TERADA, T., HANDA, N., SHIROUZU, M., AKIYAMA, T. & YOKOYAMA, S. 2011. Crystal structures of the armadillo repeat domain of adenomatous polyposis coli and its complex with the tyrosine-rich domain of Sam68. *Structure,* 19**,** 1496-508.
MOURA, I., PAULETA, S. & MOURA, J. G. 2008. Enzymatic activity mastered by altering metal coordination spheres. *JBIC Journal of Biological Inorganic Chemistry,* 13**,** 1185-1195.
MUELLER, M., WANG, M. & SCHULZE-BRIESE, C. 2012. Optimal fine phi-slicing for single-photon-counting pixel detectors. *Acta Crystallogr D Biol Crystallogr,* 68**,** 42-56.

MURPHY, M. E. P., TURLEY, S. & ADMAN, E. T. 1997. Structure of Nitrite Bound to Copper-containing Nitrite Reductase from Alcaligenes faecalis : MECHANISTIC IMPLICATIONS. *Journal of Biological Chemistry,* 272**,** 28455-28460.

MURPHY, M. E. P., TURLEY, S., KUKIMOTO, M., NISHIYAMA, M., HORINOUCHI, S., SASAKI, H., TANOKURA, M. & ADMAN, E. T. 1995. Structure of Alcaligenes faecalis Nitrite Reductase and a Copper Site Mutant, M150E, That Contains Zinc. *Biochemistry,* 34**,** 12107-12117.

MURRAY, J. & GARMAN, E. 2002. Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. *J Synchrotron Radiat,* 9**,** 347-54.

MURRAY, J. W., GARMAN, E. F. & RAVELLI, R. B. G. 2004. X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose. *Journal of Applied Crystallography,* 37**,** 513-522.

MURRAY, J. W., RUDINO-PINERA, E., OWEN, R. L., GRININGER, M., RAVELLI, R. B. & GARMAN, E. F. 2005. Parameters affecting the X-ray dose absorbed by macromolecular crystals. *J Synchrotron Radiat,* 12**,** 268-75.

MURSHUDOV, G. N., SKUBAK, P., LEBEDEV, A. A., PANNU, N. S., STEINER, R. A., NICHOLLS, R. A., WINN, M. D., LONG, F. & VAGIN, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr,* 67**,** 355-67.

NAGLE, J. F. & MOROWITZ, H. J. 1978. Molecular mechanisms for proton transport in membranes. *Proceedings of the National Academy of Sciences of the United States of America,* 75**,** 298-302.

NAKAMURA, A., WADA, C. & MIKI, K. 2007. Expression and purification of F-plasmid RepE and preliminary X-ray crystallographic study of its complex with operator DNA. *Acta Crystallographica Section F,* 63**,** 346-349.

NAVE, C. & HILL, M. A. 2005. Will reduced radiation damage occur with very small crystals? *Journal of Synchrotron Radiation,* 12**,** 299-303.

NEUTZE, R., WOUTS, R., VAN DER SPOEL, D., WECKERT, E. & HAJDU, J. 2000. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature,* 406**,** 752-7.

NEVIANI, P., SANTHANAM, R., TROTTA, R., NOTARI, M., BLASER, B. W., LIU, S., MAO, H., CHANG, J. S., GALIETTA, A., UTTAM, A., ROY, D. C., VALTIERI, M., BRUNER-KLISOVIC, R., CALIGIURI, M. A., BLOOMFIELD, C. D., MARCUCCI, G. & PERROTTI, D. 2005. The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein. *Cancer Cell,* 8**,** 355-68.

NEWMAN, J. 2006. A review of techniques for maximizing diffraction from a protein crystal in stilla. *Acta Crystallographica Section D,* 62**,** 27-31.

NISHINO, T. & OKAMOTO, K. 2000. The role of the [2Fe-2s] cluster centers in xanthine oxidoreductase. *J Inorg Biochem,* 82**,** 43-9.

NISHIYAMA, M., SUZUKI, J., OHNUKI, T., CHANG, H. C., HORINOUCHI, S., TURLEY, S., ADMAN, E. T. & BEPPU, T. 1992. Site-directed mutagenesis of pseudoazurin from Alcaligenes faecalis S-6; Pro80Ala mutant exhibits marked increase in reduction potential. *Protein Eng,* 5**,** 177-84.

NOWAK, E., BRZUSZKIEWICZ, A., DAUTER, M., DAUTER, Z. & ROSENBAUM, G. 2009. To scavenge or not to scavenge: that is the question. *Acta Crystallographica Section D,* 65**,** 1004-1006.

O'TOOLE, G. A. & KOLTER, R. 1998. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Mol Microbiol,* 30**,** 295-304.

OAKS, J. J., SANTHANAM, R., WALKER, C. J., ROOF, S., HARB, J. G., FERENCHAK, G., EISFELD, A. K., VAN BROCKLYN, J. R., BRIESEWITZ, R., SADDOUGHI, S. A., NAGATA, K., BITTMAN, R., CALIGIURI, M. A., ABDEL-WAHAB, O., LEVINE, R., ARLINGHAUS, R. B., QUINTAS-CARDAMA, A., GOLDMAN, J. M., APPERLEY, J., REID, A., MILOJKOVIC, D., ZIOLO, M. T., MARCUCCI, G., OGRETMEN, B., NEVIANI, P. & PERROTTI, D. 2013. Antagonistic activities of the immunomodulator and PP2A-activating drug FTY720 (Fingolimod, Gilenya) in Jak2-driven hematologic malignancies. *Blood,* 122**,** 1923-34.

OLESEN, K., VESELOV, A., ZHAO, Y., WANG, Y., DANNER, B., SCHOLES, C. P. & SHAPLEIGH, J. P. 1998. Spectroscopic, Kinetic, and Electrochemical Characterization of Heterologously Expressed Wild-Type and Mutant Forms of Copper-Containing Nitrite Reductase from Rhodobacter sphaeroides 2.4.3. *Biochemistry,* 37**,** 6086-6094.

OWEN, R. L., PATERSON, N., AXFORD, D., AISHIMA, J., SCHULZE-BRIESE, C., REN, J., FRY, E. E., STUART, D. I. & EVANS, G. 2014a. Exploiting fast detectors to enter a new dimension in room-temperature crystallography. *Acta Crystallogr D Biol Crystallogr,* 70**,** 1248-56.

OWEN, R. L., PATERSON, N., AXFORD, D., AISHIMA, J., SCHULZE-BRIESE, C., REN, J., FRY, E. E., STUART, D. I. & EVANS, G. 2014b. Exploiting fast detectors to enter a new dimension in room-temperature crystallography. *Acta Crystallographica Section D: Biological Crystallography,* 70**,** 1248-1256.

OWEN, R. L., RUDINO-PINERA, E. & GARMAN, E. F. 2006a. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc Natl Acad Sci U S A,* 103**,** 4912-7.

OWEN, R. L., RUDIÑO-PIÑERA, E. & GARMAN, E. F. 2006b. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 4912-4917.

PALMER, K. L., BROWN, S. A. & WHITELEY, M. 2007. Membrane-Bound Nitrate Reductase Is Required for Anaerobic Growth in Cystic Fibrosis Sputum. *Journal of Bacteriology,* 189**,** 4449-4455.

PASSADOR, L., COOK, J. M., GAMBELLO, M. J., RUST, L. & IGLEWSKI, B. H. 1993. Expression of Pseudomonas aeruginosa virulence genes requires cell-to-cell communication. *Science,* 260**,** 1127-30.

PAUL, R., ABEL, S., WASSMANN, P., BECK, A., HEERKLOTZ, H. & JENAL, U. 2007. Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. *J Biol Chem,* 282**,** 29170-7.

PAUL, R., WEISER, S., AMIOT, N. C., CHAN, C., SCHIRMER, T., GIESE, B. & JENAL, U. 2004. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes & Development,* 18**,** 715-727.

PELLEGRINI, E., PIANO, D. & BOWLER, M. W. 2011. Direct cryocooling of naked crystals: are cryoprotection agents always necessary? *Acta Crystallogr D Biol Crystallogr,* 67**,** 902-6.

PERROTTI, D., JAMIESON, C., GOLDMAN, J. & SKORSKI, T. 2010. Chronic myeloid leukemia: mechanisms of blastic transformation. *J Clin Invest,* 120**,** 2254-64.

PERROTTI, D. & NEVIANI, P. 2013. Protein phosphatase 2A: a target for anticancer therapy. *Lancet Oncol,* 14**,** e229-38.

PITT, T. L., SPARROW, M., WARNER, M. & STEFANIDOU, M. 2003. Survey of resistance of Pseudomonas aeruginosa from UK patients with cystic fibrosis to six commonly prescribed antimicrobial agents. *Thorax,* 58**,** 794-6.

POLAKIS, P. 2012. Drugging Wnt signalling in cancer. *EMBO J,* 31**,** 2737-46.

POLIKARPOV, I., TEPLYAKOV, A. & OLIVA, G. 1997. The Ultimate Wavelength for Protein Crystallography? *Acta Crystallographica Section D,* 53**,** 734-737.

POMÈS, R. & ROUX, B. 1996. Structure and dynamics of a proton wire: A theoretical study of H+ translocation along the single-file water chain in the gramicidin A channel. *Biophysical Journal,* 71**,** 19-39.

POWERS, R., COPELAND, J. C., STARK, J. L., CAPREZ, A., GURU, A. & SWANSON, D. 2011. Searching the protein structure database for ligand-binding site similarities using CPASS v.2. *BMC Res Notes,* 4**,** 17.

PRUDÊNCIO, M., EADY, R. R. & SAWERS, G. 2001. Catalytic and spectroscopic analysis of blue copper-containing nitrite reductase mutants altered in the environment of the type 2 copper centre: implications for substrate interaction. *Biochemical Journal,* 353**,** 259-266.

PULTZ, I. S., CHRISTEN, M., KULASEKARA, H. D., KENNARD, A., KULASEKARA, B. & MILLER, S. I. 2012. The response threshold of Salmonella PilZ domain proteins is determined by their binding affinities for c-di-GMP. *Mol Microbiol,* 86**,** 1424-40.

PULVERER, B. J., FISHER, C., VOUSDEN, K., LITTLEWOOD, T., EVAN, G. & WOODGETT, J. R. 1994. Site-specific modulation of c-Myc cotransformation by residues phosphorylated in vivo. *Oncogene,* 9**,** 59-70.

PUTZER, B. M. & ENGELMANN, D. 2013. E2F1 apoptosis counterattacked: evil strikes back. *Trends Mol Med,* 19**,** 89-98.

RAAD, I. 1998. Intravascular-catheter-related infections. *The Lancet,* 351**,** 893-898.

RAJENDRAN, C., DWORKOWSKI, F. S., WANG, M. & SCHULZE-BRIESE, C. 2011a. Radiation damage in room-temperature data acquisition with the PILATUS 6M pixel detector. *J Synchrotron Radiat,* 18**,** 318-28.

RAJENDRAN, C., DWORKOWSKI, F. S. N., WANG, M. & SCHULZE-BRIESE, C. 2011b. Radiation damage in room-temperature data acquisition with the PILATUS 6M pixel detector. *Journal of Synchrotron Radiation,* 18**,** 318-328.

RAMOS, G. P., ROCHA, J. L. & TUON, F. F. 2013. Seasonal humidity may influence Pseudomonas aeruginosa hospital-acquired infection rates. *Int J Infect Dis,* 17**,** e757-61.

RAO, F., YANG, Y., QI, Y. & LIANG, Z. X. 2008. Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from Pseudomonas aeruginosa. *J Bacteriol,* 190**,** 3622-31.

RAVELLI, R. B. G. & GARMAN, E. F. 2006. Radiation damage in macromolecular cryocrystallography. *Current Opinion in Structural Biology,* 16**,** 624-629.

RAVELLI, R. B. G. & MCSWEENEY, S. M. 2000. The 'fingerprint' that X-rays can leave on structures. *Structure,* 8**,** 315-328.

RILEY, M. A. 1998. Molecular mechanisms of bacteriocin evolution. *Annu Rev Genet,* 32**,** 255-78.

ROBERT-PAGANIN, J., NONIN-LECOMTE, S. & RETY, S. 2012. Crystal structure of an EAL domain in complex with reaction product 5'-pGpG. *PLoS One,* 7**,** e52424.

ROMLING, U., GALPERIN, M. Y. & GOMELSKY, M. 2013. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev,* 77**,** 1-52.

RÖMLING, U., KADER, A., SRIRAMULU, D. D., SIMM, R. & KRONVALL, G. 2005. Worldwide distribution of Pseudomonas aeruginosa clone C strains in the aquatic environment and cystic fibrosis patients. *Environmental Microbiology,* 7**,** 1029-1038.

ROSS, P., WEINHOUSE, H., ALONI, Y., MICHAELI, D., WEINBERGER-OHANA, P., MAYER, R., BRAUN, S., DE VROOM, E., VAN DER MAREL, G. A., VAN BOOM, J. H. & BENZIMAN, M. 1987. Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature,* 325**,** 279-281.

ROSSMANN, M. G. & BLOW, D. M. 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica,* 15**,** 24-31.

ROY, A. B., PETROVA, O. E. & SAUER, K. 2012. The phosphodiesterase DipA (PA5017) is essential for Pseudomonas aeruginosa biofilm dispersion. *J Bacteriol,* 194**,** 2904-15.

RUSSI, S., JUERS, D. H., SANCHEZ-WEATHERBY, J., PELLEGRINI, E., MOSSOU, E., FORSYTH, V. T., HUET, J., GOBBO, A., FELISAZ, F., MOYA, R., MCSWEENEY, S. M., CUSACK, S., CIPRIANI, F. & BOWLER, M. W. 2011. Inducing phase changes in crystals of macromolecules: Status and perspectives for controlled crystal dehydration. *Journal of Structural Biology,* 175**,** 236-243.

RUSSO KRAUSS, I., SICA, F., MATTIA, C. A. & MERLINO, A. 2012. Increasing the X-ray diffraction power of protein crystals by dehydration: the case of bovine serum albumin and a survey of literature data. *Int J Mol Sci,* 13**,** 3782-800.

RYAN, R. P., FOUHY, Y., LUCEY, J. F., CROSSMAN, L. C., SPIRO, S., HE, Y. W., ZHANG, L. H., HEEB, S., CAMARA, M., WILLIAMS, P. & DOW, J. M. 2006a. Cell-cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc Natl Acad Sci U S A,* 103**,** 6712-7.

RYAN, R. P., FOUHY, Y., LUCEY, J. F. & DOW, J. M. 2006b. Cyclic Di-GMP Signaling in Bacteria: Recent Advances and New Puzzles. *Journal of Bacteriology,* 188**,** 8327-8334.

RYAN, R. P., LUCEY, J., O'DONOVAN, K., MCCARTHY, Y., YANG, L., TOLKER-NIELSEN, T. & DOW, J. M. 2009. HD-GYP domain proteins regulate biofilm formation and virulence in Pseudomonas aeruginosa. *Environ Microbiol,* 11**,** 1126-36.

RYAN, R. P., MCCARTHY, Y., ANDRADE, M., FARAH, C. S., ARMITAGE, J. P. & DOW, J. M. 2010. Cell–cell signal-dependent dynamic interactions between HD-GYP and GGDEF domain proteins mediate virulence in Xanthomonas campestris. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 5989-5994.

RYJENKOV, D. A., TARUTINA, M., MOSKVIN, O. V. & GOMELSKY, M. 2005. Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J Bacteriol,* 187**,** 1792-8.

SAMANTA, A. K., CHAKRABORTY, S. N., WANG, Y., KANTARJIAN, H., SUN, X., HOOD, J., PERROTTI, D. & ARLINGHAUS, R. B. 2009. Jak2 inhibition deactivates Lyn kinase through the SET-PP2A-SHP1 pathway, causing apoptosis in drug-resistant cells from chronic myelogenous leukemia patients. *Oncogene,* 28**,** 1669-81.

SANCHEZ-WEATHERBY, J., BOWLER, M. W., HUET, J., GOBBO, A., FELISAZ, F., LAVAULT, B., MOYA, R., KADLEC, J., RAVELLI, R. B. & CIPRIANI, F. 2009a. Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr D Biol Crystallogr,* 65**,** 1237-46.

SANCHEZ-WEATHERBY, J., BOWLER, M. W., HUET, J., GOBBO, A., FELISAZ, F., LAVAULT, B., MOYA, R., KADLEC, J., RAVELLI, R. B. G. & CIPRIANI, F. 2009b. Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallographica Section D,* 65**,** 1237-1246.

SANISHVILI, R., NAGARAJAN, V., YODER, D., BECKER, M., XU, S., CORCORAN, S., AKEY, D. L., SMITH, J. L. & FISCHETTI, R. F. 2008. A 7mum mini-beam improves diffraction data from small or imperfect crystals of macromolecules. *Acta Crystallogr D Biol Crystallogr,* 64**,** 425-35.

SAUER, K., CAMPER, A. K., EHRLICH, G. D., COSTERTON, J. W. & DAVIES, D. G. 2002. Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm. *J Bacteriol,* 184**,** 1140-54.

SCHIRMER, T. & JENAL, U. 2009. Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol,* 7**,** 724-35.

SCHLICHTING, I., BERENDZEN, J., CHU, K., STOCK, A. M., MAVES, S. A., BENSON, D. E., SWEET, R. M., RINGE, D., PETSKO, G. A. & SLIGAR, S. G. 2000. The catalytic pathway of cytochrome p450cam at atomic resolution. *Science (New York, N.Y.),* 287**,** 1615-1622.

SCHMIDT, A. J., RYJENKOV, D. A. & GOMELSKY, M. 2005. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J Bacteriol,* 187**,** 4774-81.

SCHNEIDER, T. R. 2008. Synchrotron radiation: micrometer-sized x-ray beams as fine tools for macromolecular crystallography. *HFSP Journal,* 2**,** 302-306.

SEARS, R., NUCKOLLS, F., HAURA, E., TAYA, Y., TAMAI, K. & NEVINS, J. R. 2000. Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes Dev,* 14**,** 2501-14.

SEVILLA, M. D. & D'ARCY, J. B. 1978. An electron spin resonance investigation of charge transfer in aromatic peptide .pi.-cation radicals. *The Journal of Physical Chemistry,* 82**,** 338-342.

SEVILLA, M. D., D'ARCY, J. B. & MOREHOUSE, K. M. 1979. An electron spin resonance study of .gamma.-irradiated frozen aqueous solutions containing dipeptides. Mechanisms of radical reaction. *The Journal of Physical Chemistry,* 83**,** 2887-2892.

SHARMA, A., YEOW, W.-S., ERTEL, A., COLEMAN, I., CLEGG, N., THANGAVEL, C., MORRISSEY, C., ZHANG, X., COMSTOCK, C. E. S., WITKIEWICZ, A. K., GOMELLA, L., KNUDSEN, E. S., NELSON, P. S. & KNUDSEN, K. E. 2010. The retinoblastoma tumor suppressor controls androgen signaling and human prostate cancer progression. *The Journal of Clinical Investigation,* 120**,** 4478-4492.

SHELDRICK, G. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallographica Section D,* 66**,** 479-485.
SHIMIZU, N., HIRATA, K., HASEGAWA, K., UENO, G. & YAMAMOTO, M. 2007. Dose dependence of radiation damage for protein crystals studied at various X-ray energies. *Journal of Synchrotron Radiation,* 14**,** 4-10.
SHROUT, J. D., CHOPP, D. L., JUST, C. L., HENTZER, M., GIVSKOV, M. & PARSEK, M. R. 2006. The impact of quorum sensing and swarming motility on Pseudomonas aeruginosa biofilm formation is nutritionally conditional. *Mol Microbiol,* 62**,** 1264-77.
SJOGREN, T., CARLSSON, G., LARSSON, G., HAJDU, A., ANDERSSON, C., PETTERSSON, H. & HAJDU, J. 2002. Protein crystallography in a vapour stream: data collection, reaction initiation and intermediate trapping in naked hydrated protein crystals. *Journal of Applied Crystallography,* 35**,** 113-116.
SJOGREN, T. & HAJDU, J. 2001a. The Structure of an alternative form of Paracoccus pantotrophus cytochrome cd(1) nitrite reductase. *J Biol Chem,* 276**,** 29450-5.
SJOGREN, T. & HAJDU, J. 2001b. Structure of the bound dioxygen species in the cytochrome oxidase reaction of cytochrome cd1 nitrite reductase. *J Biol Chem,* 276**,** 13072-6.
SLABINSKI, L., JAROSZEWSKI, L., RYCHLEWSKI, L., WILSON, I. A., LESLEY, S. A. & GODZIK, A. 2007. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics,* 23**,** 3403-5.
SMITH, E. E., BUCKLEY, D. G., WU, Z., SAENPHIMMACHAK, C., HOFFMAN, L. R., D'ARGENIO, D. A., MILLER, S. I., RAMSEY, B. W., SPEERT, D. P., MOSKOWITZ, S. M., BURNS, J. L., KAUL, R. & OLSON, M. V. 2006. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proceedings of the National Academy of Sciences,* 103**,** 8487-8492.
SODING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res,* 33**,** W244-8.
SÖDING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research,* 33**,** W244-W248.
SOMMER, R., JOACHIM, I., WAGNER, S. & TITZ, A. 2013. New approaches to control infections: anti-biofilm strategies against gram-negative bacteria. *Chimia (Aarau),* 67**,** 286-90.
SOO HOO, L., ZHANG, J. Y. & CHAN, E. K. 2002. Cloning and characterization of a novel 90 kDa 'companion' auto-antigen of p62 overexpressed in cancer. *Oncogene,* 21**,** 5006-15.
STANELLE, J. & PÜTZER, B. M. 2006. E2F1-induced apoptosis: turning killers into therapeutics. *Trends in Molecular Medicine,* 12**,** 177-185.
STARKENBURG, S. R., ARP, D. J. & BOTTOMLEY, P. J. 2008. Expression of a putative nitrite reductase and the reversible inhibition of nitrite-dependent respiration by nitric oxide in Nitrobacter winogradskyi Nb-255. *Environ Microbiol,* 10**,** 3036-42.
STEIN, N. 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography,* 41**,** 641-643.
STICKLER, D., MORRIS, N., MORENO, M. C. & SABBUBA, N. 1998. Studies on the formation of crystalline bacterial biofilms on urethral catheters. *Eur J Clin Microbiol Infect Dis,* 17**,** 649-52.
STRANGE, R. W., DODD, F. E., ABRAHAM, Z. H. L., GROSSMANN, J. G., BRUSER, T., EADY, R. R., SMITH, B. E. & HASNAIN, S. S. 1995. The substrate-binding site in Cu nitrite reductase and its similarity to Zn carbonic anhydrase. *Nat Struct Mol Biol,* 2**,** 287-292.
STRANGE, R. W., MURPHY, L. M., DODD, F. E., ABRAHAM, Z. H. L., EADY, R. R., SMITH, B. E. & HASNAIN, S. S. 1999. Structural and kinetic evidence for an ordered mechanism of copper nitrite reductase1. *Journal of Molecular Biology,* 287**,** 1001-1009.
STUDENT 1908. The Probable Error of a mean. *Biometrika* 6**,** 25.
SUDARSAN, N., LEE, E. R., WEINBERG, Z., MOY, R. H., KIM, J. N., LINK, K. H. & BREAKER, R. R. 2008. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science,* 321**,** 411-3.
SUH, J. 1992. Model studies of metalloenzymes involving metal ions as Lewis acid catalysts. *Accounts of Chemical Research,* 25**,** 273-279.

SULTAN, S. Z., PITZER, J. E., BOQUOI, T., HOBBS, G., MILLER, M. R. & MOTALEB, M. A. 2011. Analysis of the HD-GYP domain cyclic dimeric GMP phosphodiesterase reveals a role in motility and the enzootic life cycle of Borrelia burgdorferi. *Infect Immun,* 79**,** 3273-83.

SUNDRIYAL, A., MASSA, C., SAMORAY, D., ZEHENDER, F., SHARPE, T., JENAL, U. & SCHIRMER, T. 2014. Inherent regulation of EAL domain-catalyzed hydrolysis of second messenger cyclic di-GMP. *J Biol Chem,* 289**,** 6978-90.

SUZUKI, S., KATAOKA, K. & YAMAGUCHI, K. 2000. Metal Coordination and Mechanism of Multicopper Nitrite Reductase. *Accounts of Chemical Research,* 33**,** 728-735.

TAKAYAMA, Y. & NAKASAKO, M. 2011. A few low-frequency normal modes predominantly contribute to conformational responses of hen egg white lysozyme in the tetragonal crystal to variations of molecular packing controlled by environmental humidity. *Biophys Chem,* 159**,** 237-46.

TAMAYO, R., TISCHLER, A. D. & CAMILLI, A. 2005. The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *J Biol Chem,* 280**,** 33324-30.

TARNAWSKI, M., BARENDS, T. R., HARTMANN, E. & SCHLICHTING, I. 2013. Structures of the catalytic EAL domain of the Escherichia coli direct oxygen sensor. *Acta Crystallogr D Biol Crystallogr,* 69**,** 1045-53.

TAYLOR, G. L. 2010. Introduction to phasing. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 325-338.

TCHIGVINTSEV, A., XU, X., SINGER, A., CHANG, C., BROWN, G., PROUDFOOT, M., CUI, H., FLICK, R., ANDERSON, W. F., JOACHIMIAK, A., GALPERIN, M. Y., SAVCHENKO, A. & YAKUNIN, A. F. 2010. Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. *J Mol Biol,* 402**,** 524-38.

TENG, T.-Y. & MOFFAT, K. 2000. Primary radiation damage of protein crystals by an intense synchrotron X-ray beam. *Journal of Synchrotron Radiation,* 7**,** 313-317.

TEPLYAKOV, A., OLIVA, G. & POLIKARPOV, I. 1998. On the choice of an optimal wavelength in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr,* 54**,** 610-4.

TESMER, J. J., SUNAHARA, R. K., JOHNSON, R. A., GOSSELIN, G., GILMAN, A. G. & SPRANG, S. R. 1999. Two-metal-Ion catalysis in adenylyl cyclase. *Science,* 285**,** 756-60.

THODEN, J. B. & HOLDEN, H. M. 2007. The molecular architecture of glucose-1-phosphate uridylyltransferase. *Protein Science : A Publication of the Protein Society,* 16**,** 432-440.

THORMANN, K. M., DUTTLER, S., SAVILLE, R. M., HYODO, M., SHUKLA, S., HAYAKAWA, Y. & SPORMANN, A. M. 2006. Control of Formation and Cellular Detachment from Shewanella oneidensis MR-1 Biofilms by Cyclic di-GMP. *Journal of Bacteriology,* 188**,** 2681-2691.

THORNE, R. E., STUM, Z., KMETKO, J., O'NEILL, K. & GILLILAN, R. 2003. Microfabricated mounts for high-throughput macromolecular cryocrystallography. *Journal of Applied Crystallography,* 36**,** 1455-1460.

TOCHEVA, E. I., ROSELL, F. I., MAUK, A. G. & MURPHY, M. E. 2004. Side-on copper-nitrosyl coordination by nitrite reductase. *Science,* 304**,** 867-70.

TOLKER-NIELSEN, T. 2014. Pseudomonas aeruginosa biofilm infections: From molecular biofilm biology to new treatment possibilities. *APMIS,* 122**,** 1-51.

TRIGG, J., GUTWIN, K., KEATING, A. E. & BERGER, B. 2011. Multicoil2: Predicting Coiled Coils and Their Oligomerization States from Sequence in the Twilight Zone. *PLoS ONE,* 6**,** e23519.

TSOLIS, A. C., PAPANDREOU, N. C., ICONOMIDOU, V. A. & HAMODRAKAS, S. J. 2013. A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS One,* 8**,** e54175.

TYCKO, R. 2004. Progress towards a molecular-level structural understanding of amyloid fibrils. *Curr Opin Struct Biol,* 14**,** 96-103.

VAGIN, A. & TEPLYAKOV, A. 2010a. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr,* 66**,** 22-5.

VAGIN, A. & TEPLYAKOV, A. 2010b. Molecular replacement with MOLREP. *Acta Crystallographica Section D,* 66**,** 22-25.
VAIDYANATHAN, S., SONI, B. M., OO, T., HUGHES, P. L., MANSOUR, P. & SINGH, G. 2009. Infection of Brindley sacral anterior root stimulator by Pseudomonas aeruginosa requiring removal of the implant: long-term deleterious effects on bowel and urinary bladder function in a spinal cord injury patient with tetraplegia: a case report. *Cases J,* 2**,** 9364.
VALLET, I., OLSON, J. W., LORY, S., LAZDUNSKI, A. & FILLOUX, A. 2001. The chaperone/usher pathways of Pseudomonas aeruginosa: Identification of fimbrial gene clusters (cup) and their involvement in biofilm formation. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 6911-6916.
VANDER WAUVEN, C., PIÉRARD, A., KLEY-RAYMANN, M. & HAAS, D. 1984. Pseudomonas aeruginosa mutants affected in anaerobic growth on arginine: evidence for a four-gene cluster encoding the arginine deiminase pathway. *Journal of Bacteriology,* 160**,** 928-934.
VENTELA, S., COME, C., MAKELA, J. A., HOBBS, R. M., MANNERMAA, L., KALLAJOKI, M., CHAN, E. K., PANDOLFI, P. P., TOPPARI, J. & WESTERMARCK, J. 2012. CIP2A promotes proliferation of spermatogonial progenitor cells and spermatogenesis in mice. *PLoS One,* 7**,** e33209.
VERHOEFF, M., VAN DER VEEN, E. L., ROVERS, M. M., SANDERS, E. A. M. & SCHILDER, A. G. M. 2006. Chronic suppurative otitis media: A review. *International Journal of Pediatric Otorhinolaryngology,* 70**,** 1-12.
VIJAYALAKSHMI, L., KRISHNA, R., SANKARANARAYANAN, R. & VIJAYAN, M. 2008. An asymmetric dimer of β-lactoglobulin in a low humidity crystal form—Structural changes that accompany partial dehydration and protein action. *Proteins: Structure, Function, and Bioinformatics,* 71**,** 241-249.
VINCENT, T. L., GREEN, P. J. & WOOLFSON, D. N. 2013. LOGICOIL--multi-state prediction of coiled-coil oligomeric state. *Bioinformatics,* 29**,** 69-76.
VON SONNTAG, C. & SCHUCHMANN, H.-P. 1994. Suppression of hydroxyl radical reactions in biological systems: Considerations based on competition kinetics. *In:* LESTER, P. (ed.) *Methods in Enzymology.* Academic Press.
WALLS, D. & LOUGHRAN, S. T. 2011. Tagging recombinant proteins to enhance solubility and aid purification. *Methods Mol Biol,* 681**,** 151-75.
WANG, B. C. 1985. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol,* 115**,** 90-112.
WANG, X., HILLS, L. B. & HUANG, Y. H. 2015. Lipid and Protein Co-Regulation of PI3K Effectors Akt and Itk in Lymphocytes. *Front Immunol,* 6**,** 117.
WARKENTIN, M. & THORNE, R. E. 2009. Slow cooling of protein crystals. *J Appl Crystallogr,* 42**,** 944-952.
WARREN, A. J., ARMOUR, W., AXFORD, D., BASHAM, M., CONNOLLEY, T., HALL, D. R., HORRELL, S., MCAULEY, K. E., MYKHAYLYK, V., WAGNER, A. & EVANS, G. 2013. Visualization of membrane protein crystals in lipid cubic phase using X-ray imaging. *Acta Crystallogr D Biol Crystallogr,* 69**,** 1252-9.
WASSER, I. M., DE VRIES, S., MOENNE-LOCCOZ, P., SCHRODER, I. & KARLIN, K. D. 2002. Nitric oxide in biological denitrification: Fe/Cu metalloenzyme and metal complex NO(x) redox chemistry. *Chem Rev,* 102**,** 1201-34.
WASSMANN, P., CHAN, C., PAUL, R., BECK, A., HEERKLOTZ, H., JENAL, U. & SCHIRMER, T. 2007. Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure,* 15**,** 915-27.
WATANABE, N., WATANABE, N., WATANABE, N., WATANABE, N. & WATANABE, N. 2006. From phasing to structure refinement in-house: Cr/Cu dual-wavelength system and a loopless free crystal-mounting method. *Acta Crystallographica Section D: Biological Crystallography,* 62**,** 891.

WEICHMAN, K. E., LEVINE, S. M., WILSON, S. C., CHOI, M. & KARP, N. S. 2013. Antibiotic selection for the treatment of infectious complications of implant-based breast reconstruction. *Ann Plast Surg,* 71**,** 140-3.

WEIK, M., RAVELLI, R. B. G., KRYGER, G., MCSWEENEY, S., RAVES, M. L., HAREL, M., GROS, P., SILMAN, I., KROON, J. & SUSSMAN, J. L. 2000. Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proceedings of the National Academy of Sciences,* 97**,** 623-628.

WEINERT, T., OLIERIC, V., WALTERSPERGER, S., PANEPUCCI, E., CHEN, L., ZHANG, H., ZHOU, D., ROSE, J., EBIHARA, A., KURAMITSU, S., LI, D., HOWE, N., SCHNAPP, G., PAUTSCH, A., BARGSTEN, K., PROTA, A. E., SURANA, P., KOTTUR, J., NAIR, D. T., BASILICO, F., CECATIELLO, V., PASQUALATO, S., BOLAND, A., WEICHENRIEDER, O., WANG, B.-C., STEINMETZ, M. O., CAFFREY, M. & WANG, M. 2015. Fast native-SAD phasing for routine macromolecular structure determination. *Nat Meth,* 12**,** 131-133.

WEISS, M. 2001. Global indicators of X-ray data quality. *Journal of Applied Crystallography,* 34**,** 130-135.

WEISS, M. S. & HILGENFELD, R. 1997. On the use of the merging R factor as a quality indicator for X-ray data. *Journal of Applied Crystallography,* 30**,** 203-205.

WESTERMARCK, J. & HAHN, W. C. 2008. Multiple pathways regulated by the tumor suppressor PP2A in transformation. *Trends Mol Med,* 14**,** 152-60.

WHALLEY, K. L. 1973. *Properties of trapped and mobile electrons in pure ice and in alkaline ice,* United States, Royal Society of Canada.

WHEELER, M. J., RUSSI, S., BOWLER, M. G. & BOWLER, M. W. 2012. Measurement of the equilibrium relative humidity for common precipitant concentrations: facilitating controlled dehydration experiments. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 68**,** 111-4.

WHITCHURCH, C. B., TOLKER-NIELSEN, T., RAGAS, P. C. & MATTICK, J. S. 2002. Extracellular DNA Required for Bacterial Biofilm Formation. *Science,* 295**,** 1487.

WIERMAN, J. L., ALDEN, J. S., KIM, C. U., MCEUEN, P. L., GRUNER, S. M., WIERMAN, J. L., ALDEN, J. S., KIM, C. U., MCEUEN, P. L. & GRUNER, S. M. 2013. Graphene as a protein crystal mounting material to reduce background scatter. *Journal of Applied Crystallography,* 46**,** 1501.

WIJMA, H. J., JEUKEN, L. J. C., VERBEET, M. P., ARMSTRONG, F. A. & CANTERS, G. W. 2006. A Random-sequential Mechanism for Nitrite Binding and Active Site Reduction in Copper-containing Nitrite Reductase. *Journal of Biological Chemistry,* 281**,** 16340-16346.

WINKLER, A., UDVARHELYI, A., HARTMANN, E., REINSTEIN, J., MENZEL, A., SHOEMAN, R. L. & SCHLICHTING, I. 2014. Characterization of elements involved in allosteric light regulation of phosphodiesterase activity by comparison of different functional BlrP1 states. *J Mol Biol,* 426**,** 853-68.

WINSTANLEY, C., LANGILLE, M. G. I., FOTHERGILL, J. L., KUKAVICA-IBRULJ, I., PARADIS-BLEAU, C., SANSCHAGRIN, F., THOMSON, N. R., WINSOR, G. L., QUAIL, M. A., LENNARD, N., BIGNELL, A., CLARKE, L., SEEGER, K., SAUNDERS, D., HARRIS, D., PARKHILL, J., HANCOCK, R. E. W., BRINKMAN, F. S. L. & LEVESQUE, R. C. 2009. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome Research,* 19**,** 12-23.

WINTER, G., LOBLEY, C. M. & PRINCE, S. M. 2013. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr,* 69**,** 1260-73.

XUE, Y., WU, G., WANG, X., ZOU, X., ZHANG, G., XIAO, R., YUAN, Y., LONG, D., YANG, J., WU, Y., XU, H., LIU, F. & LIU, M. 2013. CIP2A is a predictor of survival and a novel therapeutic target in bladder urothelial cell carcinoma. *Med Oncol,* 30**,** 406.

YAGER, D. R. & NWOMEH, B. C. 1999. The proteolytic environment of chronic wounds. *Wound Repair Regen,* 7**,** 433-41.

YANG, L., NILSSON, M., GJERMANSEN, M., GIVSKOV, M. & TOLKER-NIELSEN, T. 2009. Pyoverdine and PQS mediated subpopulation interactions involved in Pseudomonas aeruginosa biofilm formation. *Mol Microbiol,* 74**,** 1380-92.

YANO, J., KERN, J., IRRGANG, K.-D., LATIMER, M. J., BERGMANN, U., GLATZEL, P., PUSHKAR, Y., BIESIADKA, J., LOLL, B., SAUER, K., MESSINGER, J., ZOUNI, A. & YACHANDRA, V. K. 2005a. X-ray damage to the Mn4Ca complex in single crystals of photosystem II: A case study for metalloprotein crystallography. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 12047-12052.

YANO, J., KERN, J., IRRGANG, K. D., LATIMER, M. J., BERGMANN, U., GLATZEL, P., PUSHKAR, Y., BIESIADKA, J., LOLL, B., SAUER, K., MESSINGER, J., ZOUNI, A. & YACHANDRA, V. K. 2005b. X-ray damage to the Mn4Ca complex in single crystals of photosystem II: a case study for metalloprotein crystallography. *Proc Natl Acad Sci U S A,* 102**,** 12047-52.

YAO, D., HUANG, S., WANG, J., GU, Y., ZHENG, C. & FAN, H. 2005. SAD phasing at Bijvoet ratio below 0.6%. *Acta Crystallographica Section A,* 61**,** c154.

YAP, T. L., CHEN, Y. L., XU, T., WEN, D., VASUDEVAN, S. G. & LESCAR, J. 2007. A multi-step strategy to obtain crystals of the dengue virus RNA-dependent RNA polymerase that diffract to high resolution. *Acta Crystallographica Section F,* 63**,** 78-83.

YEH, E., CUNNINGHAM, M., ARNOLD, H., CHASSE, D., MONTEITH, T., IVALDI, G., HAHN, W. C., STUKENBERG, P. T., SHENOLIKAR, S., UCHIDA, T., COUNTER, C. M., NEVINS, J. R., MEANS, A. R. & SEARS, R. 2004. A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat Cell Biol,* 6**,** 308-18.

ZELDIN, O. B., GERSTEL, M. & GARMAN, E. F. 2013. RADDOSE-3D: time- and space-resolved modelling of dose in macromolecular crystallography. *Journal of Applied Crystallography,* 46**,** 1225-1230.

ZETH, K., ROMER, C., PATZER, S. I. & BRAUN, V. 2008. Crystal structure of colicin M, a novel phosphatase specifically imported by Escherichia coli. *J Biol Chem,* 283**,** 25324-31.

ZHANG, Z., LOUBOUTIN, J. P., WEINER, D. J., GOLDBERG, J. B. & WILSON, J. M. 2005. Human airway epithelial cells sense Pseudomonas aeruginosa infection via recognition of flagellin by Toll-like receptor 5. *Infect Immun,* 73**,** 7151-60.

ZUMFT, W. G. 1997. Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews,* 61**,** 533-616.